UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

URBAN TRAVEL DEMAND : THE IMPACT OF BOX-COX TRANSFORMATIONS WITH NONSPHERICAL RESIDUAL ERRORS

by

Marcel G. Dagenais
Marc J.I. Gaudry
Tran Cong Liem

The research for this paper was supported at various stages by Transport Canada, the F.C.A.C. program of the ministère de l'Education du Québec, the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council of Canada. A first and widely circulated version, "Multiple Regression Analysis with Box-Cox Transformations : a Transportation Application", C.R.T., Publication # 166 (1980), contained partial results on one of the three equations analysed here. Early estimates were presented at the June 1979 meeting of the Econometric Society in Montreal. Lina and Lisa DiPietro made significant contributions as research assistants. A part of Gaudry's contribution was made during his tenure as an Alexander von Humboldt Research Scholar in 1985 at Universität Karlsruhe where Reiner Koblo provided computing assistance.

Institut für Wirtschaftspolitik und Wirtschaftsforschung
Discussion paper 6/85
Universität Karlsruhe

Centre de recherche sur les transports - Publication # 358
Département de sciences économiques - Cahier # 8431
Université de Montréal
C.P. 6128, Succursale A, Montréal, Canada H3C 3J7

June 1985

Dépôt légal - 4e trimestre 1985

CONTENTS

Abstract.....iv
List of figures and tables.....iii

1. Antecedents and state of the question.....1
    1.1. Starting point: The DEMTEC direct demand model.....1
    1.2. Introducing Box-Cox transformations (BCT) on the systematic part....4
    1.3. Adding a flexible form for heteroskedasticity (HG).....9
    1.4. The practical specification of the expanded model.....11

2. The likelihood function.....14
    2.1. A two-limit Tobit approach.....14
    2.2. In the absence of limit observations.....17
    2.3. The expected value of the dependent variable and its probability of being at the limits.....20
        2.3.1. Limiting behavior of the expected value 21
        2.3.2. Probability that the sample contains limit observations....24

3. The L-1.1 algorithm.....25
    3.1. The concentrated likelihood.....26
    3.2. Maximization and determination of autoregressive structure (AU)....27
    3.3. t-Stat.: unconditional but not invariant to units of measurement...28
    3.4. Analogs of the multiple correlation coefficient R^2.....30

4. Results: functional form, autocorrelation and heteroscedasticity.....32
    4.1. A table of likelihood and BCT values.....32
    4.2. General comments.....36
        4.2.1. Reasonableness of the stochastic approach.....36
        4.2.2. Behavior of the likelihood function.....36
    4.3. Flexible format questions.....37
        4.3.1. The usefulness of BCT on the systematic part.....37
        4.3.2. A general form for heteroscedasticity.....39
        4.3.3. A multiple autocorrelation structure.....40
            - the Box-Cox "arts".....40
            - a conflict of tests?.....41
    4.4. Interaction questions.....41
        4.4.1. The influence of AU and HG on BCT.....42
        4.4.2. The influence of HG on AU.....43
        4.4.3. The influence of BCT on AU and HG.....43

5. Results: elasticities and values of time.....44
    5.1. Intuitive versus strict elasticity.....44
    5.2. Elasticity with respect to variables which contain observed zeros..47
        5.2.1. Quasi-dummy variables.....47
        5.2.2. Dummy variables.....49
        5.2.3. Approximation used and reporting format.....54
    5.3. Elasticities.....56
        5.3.1. The three demand equations.....58
        5.3.2. Pushing the model to its limits.....61
    5.4. Values of time.....62

6. Conclusion.....66

Appendix: on the t-Statistics of the regression coefficients.....68
Acknowledgements.....70
References.....71
Footnotes.....76

LIST OF FIGURES AND TABLES

Figure 1. Influence of a logarithmic transformation on the fit and error variance.....8

Figure 2. Dimensions and their interactions established from Table 1.....38

Table 1. Log likelihood values (L) and Box-Cox lambda parameters (BCT) obtained with the L-1.1 algorithm on the DEMTEC model of monthly demand for transit (adults, schoolchildren) and on a model of monthly demand for gasoline in Montreal.....33

Table 2. Special cases of expected value elasticity for binary variables....53

Table 3. Elasticities and other statistics: selected cases from series A-1, S-1 and G-1 of Table 1.....59

Table 4. Elasticities and other statistics: selected cases from series S-4 of Table 1.....63

Table 5. Implicit values of time, adults and schoolchildren.....65

Table 6. Conditional and unconditional t-Statistics with different scaling rules.....69

ABSTRACT

We develop an algorithm for solving regression models with Box-Cox transformations on both the dependent and independent variables, while simultaneously taking into account corrections for serial correlation of several orders and for heteroscedasticity. The latter correction is of a general form which contains as special cases most specifications of heteroscedasticity found in practice. We apply the procedure to three urban travel demand functions, two of which are currently used in their linear form by the Montreal Transit Authority and analyse more than 100 specifications. Our results show that taking into account nonsphericalness of the residuals has a major impact on model parameter estimates, notably on those which determine the functional form of the model, and that, conversely, modifications of the functional form have strong implications for both the structure of autocorrelation and the importance of heteroscedasticity; moreover, we find interactions between autocorrelation and heteroscedasticity structures. We introduce a special measure of elasticity for variables which contain zero observations, particularly dummy variables. Moreover, we find that elasticities of demand and implicit values of time depend to a large extent on the stochastic specification of the model.

KEY WORDS: TRANSIT DEMAND, GASOLINE DEMAND, BOX-COX TRANSFORMATIONS, AUTOCORRELATION, HETEROSCEDASTICITY, ELASTICITIES FOR BINARY VARIABLES, VALUE OF TIME

RESUME

Nous développons un algorithme pour résoudre des modèles de régression comportant des transformations de Box-Cox sur les variables dépendante et indépendantes en tenant simultanément compte de l'autocorrélation multiple et de l'hétéroscedasticité présentes dans les résidus. La correction faite pour tenir compte de l'hétéroscédasticité est d'une forme très générale qui comprend comme cas particuliers la plupart des formulations utilisées en pratique. Nous appliquons notre procédure à trois fonctions de demande de transport de voyageurs en milieu urbain montréalais, dont deux sont couramment utilisées en version linéaire par la Commission de Transport de la ville pour l'analyse et la prévision de la demande de transport en commun. Nos résultats, fondés sur l'analyse de plus d'une centaine de formulations, montrent à l'évidence l'interdépendance entre les paramètres qui déterminent la forme fonctionnelle de la partie systématique d'un modèle et ceux qui décrivent sa partie aléatoire. Nous introduisons une mesure d'élasticité particulière pour les variables comportant des observations nulles, en particulier les variables auxiliaires. Nous trouvons par ailleurs que les élasticités de la demande et les prix implicites du temps sont sensibles à la spécification stochastique retenue pour la partie aléatoire du modèle.

MOTS-CLES: DEMANDE DE TRANSPORT EN COMMUN, DEMANDE D'ESSENCE, TRANSFORMATIONS BOX-COX, AUTOCORRELATION, HETEROSCEDASTICITE, ELASTICITE PAR RAPPORT À UNE VARIABLE BINAIRE, VALEUR DU TEMPS

1. ANTECEDENTS AND STATE OF THE QUESTION

This paper formulates and extensively applies an algorithm (L-1.1) to estimate the parameters of an econometric specification structure which generalizes the Box-Cox regression method by simultaneously accounting for multiple autocorrelation and heteroscedasticity of a broad and flexible nature. The applications are made with two transit demand equations used in practice as well as with a gasoline demand equation and summarize a large number of tests and considerable numerical experience with the procedure. In addition to the statistical results, we present economic results through the use of tables of rigourously defined elasticities, including new measures for dummy and quasi-dummy variables, and values of time.

The best way to lead into such a fare is to summarize the antecedents of the matter, starting with the simple form of the model used for our tests, and to gradually motivate the buildup of the econometric specification.

1.1. Starting Point : the DEMTEC Direct Demand Model

One of the authors has presented in this Journal an aggregate monthly time-series demand model where all of the variables are defined over the whole urban area served by the transit authority (Gaudry, 1975, 1978). For each market category c, we write, for each time period t, the demand d_{ct} as :

$$d_{ct} = d\ (P_t, M_t, N_t, Y_t, A_t, ETC_t\ ;\ \theta_c\ ;\ w_{ct}) \quad , \qquad (1)$$

where the transit demand is explained by a set of prices P, the availability of cars M, network service levels and infrastructure characteristics N, income Y, activities which generate travel A, and data-source specific variables ETC. The symbol θ denotes the set of all parameters to be estimated and w_{ct} an error term. The two market categories analysed were adults and schoolchildren (less than 18 years of age). Later, formulation (1) was used to specify a demand for gasoline equation in the same urban market and the model was extended to include supply and network performance/cost equations (Gaudry, 1980); the equations were studied both one at the time and as a simultaneous system.

When equations were analysed one by one, their econometric specification was :

$$y_t = \sum_{k=1}^{K} \beta_k X_{kt} + u_t \qquad (2a)$$

$$u_t = \psi v_t \ , \qquad \psi = 1 \ , \qquad (3a)$$

$$v_t = \sum_{\ell} \rho_\ell v_{t-\ell} + w_t \ . \qquad (4)$$

In this formulation, the "systematic" part (2a) is assumed to be linear. The stochastic part is two-sided : first, (3a) states that there is no function of explanatory variables associated with the fluctuations of u_t : the ψ is a constant set equal to 1. In addition to this assumption of homoscedasticity, we make in (4) the assumption that there is a stationary autoregressive structure with w_t, a true white noise of zero mean and constant

variance σ^2. In practice Box-Jenkins (1970) techniques are used to identify such an autoregressive structure and significant moving average components which the series v_t might have contained are approximated by increasing the number of non-zero orders of the autoregressive process. With a fixed functional form and the assumption of homoscedasticity, this procedure appears to function well. The systematic part contains the "meaningful" variables and includes the activity level variables which insure the stationarity of u_t; furthermore, an appropriate multiple autocorrelation scheme removes any predictable component of information left in the residuals over time. We have regularly found that using that information improves the regression results : this is not altogether surprising if autoregressive structures compensate for missing variables in (2a), but very hard to prove in nontrivial ways. The combination of (2a) and (4) yields a balanced specification where both systematic and stochastic parts have meaningful roles : the former includes "causal" economic and policy variables of interest and the latter whitewashes the residuals without incurring the extra computational costs associated with the estimation of moving average parameters.

The Montreal Urban Community Transit Commission (MUCTC) has been using the two transit demand functions with the above econometric specification since 1975. The model, called DEMTEC, works very well, tracking ridership to within about 1% when the level of the explanatory variables are known. Forecasts of the impact of new subway lines, of fare and service level changes and of the contributions of car ownership or economic activities

are made and updated at regular intervals. After an extensive examination of models of this sort, the Toronto Transit Commission (TTC) has implemented the procedure. The Paris Transit Authority (RATP) is considering the same action.

One might then ask : can one do better by relaxing some of the assumptions built into structure (2a, 3a, 4) ?

1.2. Introducing Box-Cox Transformations (BCT) on the Systematic Part

The first improvement one might consider is the choice of the appropriate functional form, which is a crucial step in applied regression analysis : Theil (1971) has pointed out the importance of discriminating between linear and log-linear models for forecasting. A useful compromise has been suggested by Box and Cox (196.) : applied both to the dependent and independent variables, the Box-Cox transformation (BCT) defines a general class of models which includes the pure linear and log-linear models as special cases. The flexibility of the BCT had made it attractive in monetary economics [Zarembka (1968), Spitzer (1976, 1977)] and income analysis [Heckman and Polachek (1974)] and it was natural to apply it in transportation. The following was defined as a substitute for (2a) :

$$y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t \quad , \tag{2b}$$

where the superscripts (λ_y) and (λ_{xk}) denote that the variables are affected by the BCT, for instance

$$X_{kt}^{(\lambda_{xk})} = \frac{X^{\lambda_{xk}} - 1}{\lambda_{xk}} \quad . \tag{5}$$

Clearly, BCT are applied only to strictly positive variables; other variables are left untransformed. Also, since BCT on y_t are defined only for y_t greater than zero, it is evident that equation (2b) cannot hold for every conceivable value of w_t. This point will be further discussed in the next Section [see especially equations (15) - (17)].

The application made in Gaudry and Wills (1978) used only one λ parameter, the same on both the dependent and independent variables. Both equations of the linear DEMTEC model were analysed. Their linearity could not be rejected[1]. It was also clear that if a logarithmic model had been chosen, some of the explanatory variables, such as travel time in the schoolchildren equation, would have had unexpected signs and that, in both equations, income and some of the autoregressive parameters would have had signs which differed from those obtained with the optimal functional form.

Of course, the application of the BCT, particularly to the dependent variable of regression models, does not merely affect the functional form of the model. It also affects the distributional properties of the residual errors. For instance, Savin and White (1978) found an interaction between first order autocorrelation and the BCT. In the above-mentioned application to the DEMTEC equations, where multiple autocorrelation schemes were used, the autocorrelation structure which whitewashed the residuals was the same at the

---

1. The BCT estimates were 0.92 for the adult market and 0.72 for the schoolchildren market. These values are taken from below in Table 1, Experiment # 2.1, Series A-1 and S-1; they differ slightly from the less precise estimates shown on Figures 9 and 15 of the Gaudry and Wills (1978) paper.

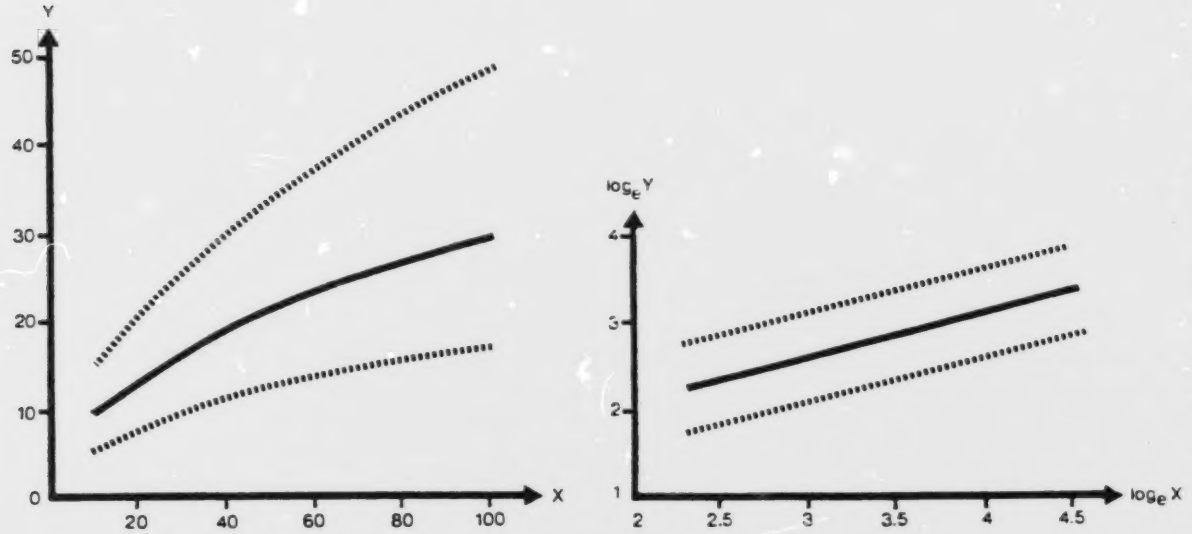
(near-linear) point of the optimal $\tilde{\lambda}$ as at the linear starting point and no tests were made to detect (by finding what the forms would have been without taking serial correlation into account) the influence of the presence of the autocorrelation structure on the optimal form of the systematic part. In the same paper, it was also felt appropriate to look for evidence of sensitivity to heteroscedasticity. Box and Cox (1964) mention that the use of the BCT may render the density of the error term more normal-like; Zarembka (1974) points out that using the BCT may compensate for heteroscedasticity and suggests an approximate measure of sensitivity conditional upon a belief in the degree of heteroscedasticity present. Using that approximate conditional procedure, it was found that, if one strongly believed in the presence of heteroscedasticity, the BCT estimates for both equations were sensitive to it and might have reflected its presence.

Clearly, this p nt is important and requires attention, for nothing guarantees that using the BCT may decrease heteroscedasticity : the BCT transformation simultaneously affects the functional form of the systematic part of the regression and the form of the density of the residual errors —and this can have disadvantages. To see this, consider the three cases presented in Figure 1 : on the left hand side, one has "initial" models; these are the object of logarithmic transformations on the right hand side. The initial models are generated by the following formulas :

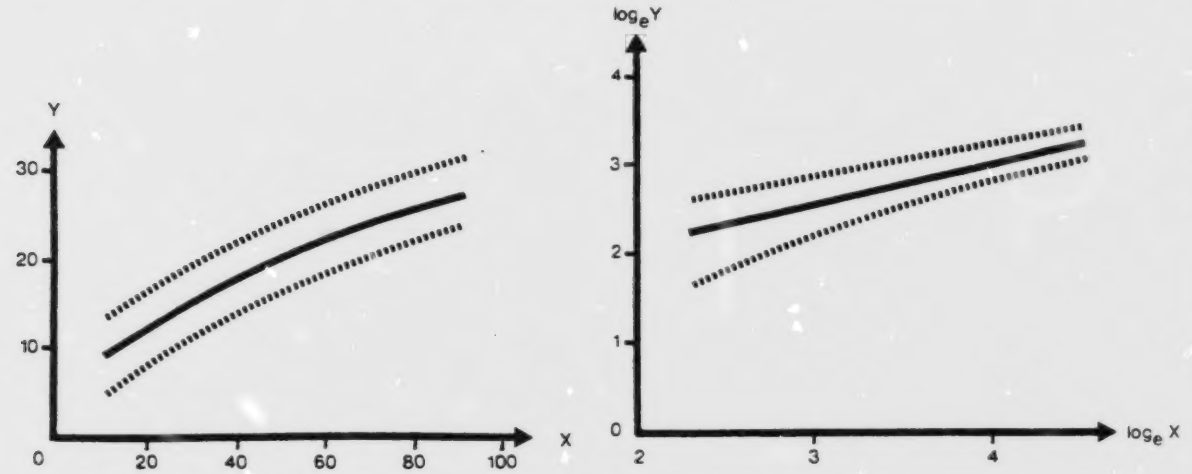
$$\left.\begin{aligned}
&\text{Case A :} \quad y_a = \beta_1 X^{\beta_2} + (y_a - \beta_1 X^{\beta_2}) = \beta_1 X^{\beta_2} + u_a \\
&\text{Case B :} \quad y_b = \beta_1 X^{\beta_2} + (y_b - \beta_1 X^{\beta_2}) = \beta_1 X^{\beta_2} + u_b \\
&\text{Case C :} \quad y_c = \beta_1 + \beta_2 X + [y_c - (\beta_1 + \beta_2 X)] = \beta_1 + \beta_2 X + u_c \quad .
\end{aligned}\right\} \quad (6)$$

Case A presents the nice situation in which a curvilinear model with heteroscedastic disturbances is transformed to obtain a better fit which simultaneously reduces the heteroscedasticity. In Case B, however, the curvilinear model has homoscedastic disturbances and the transformation introduces undesired heteroscedasticity. Alternatively, Case C presents a linear model with heteroscedastic disturbances where the use of the BCT reduces the heteroscedasticity but impairs the linear fit. In order to avoid such a dilemma, one should consider separate instruments for the different targets : the BCT to improve the functional fit and a different scheme to correct for heteroscedasticity. This strategy, suggested in Gaudry and Dagenais (1978, 1979) will be substantiated here at the same time as the generality of equation (2b) will be exploited —all without neglecting the multiple autocorrelation structure (4). Indeed, the autocorrelation structure can also be influenced by the presence of the BCT and heteroscedasticity : it requires simultaneous estimation.

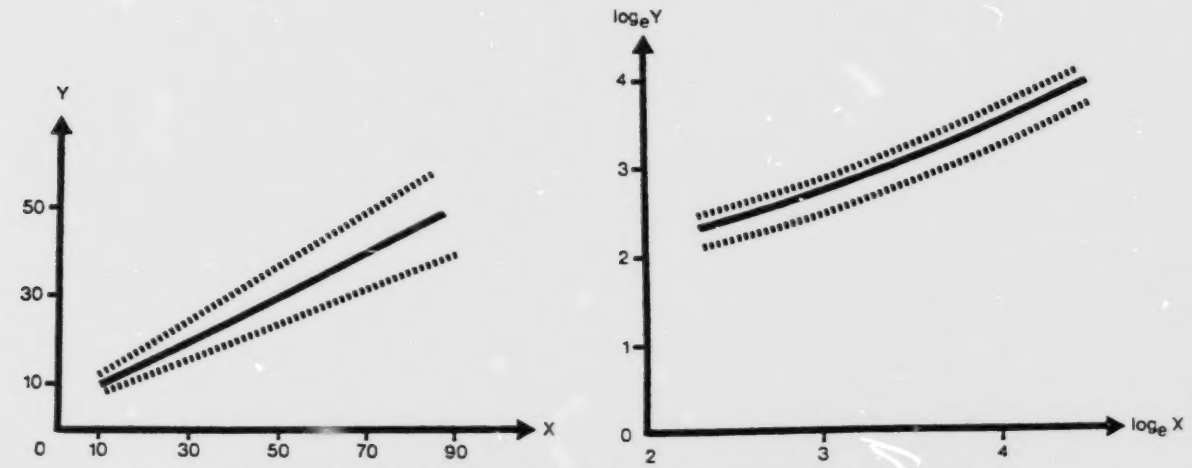
Although the BCT has continued to be used by Welland (1978) in income analysis, by Appelbaum (1979) and by Berndt and Khaled (1979) in production theory, and by other authors, the specification of the stochastic part has been



Case A. Reducing both nonlinearity and heteroscedasticity



Case B. Reducing nonlinearity but inducing heteroscedasticity



Case C. Reducing heteroscedasticity but inducing nonlinearity

INFLUENCE OF A LOGARITHMIC TRANSFORMATION ON THE FIT AND ERROR VARIANCE

FIGURE 1

generally neglected. In transportation, for instance, Kau and Sirmans (1976) have a simple gravity model of interregional migration, Wang et al. (1981), Greene (1982) and Wang and Skinner (1984) have time-series models of freight flows, gasoline consumption and transit demand, respectively, but none take autocorrelation or heteroscedasticity into account. Neither have we found elsewhere any examination of the interaction among functional form, heteroscedasticity and autocorrelation which this paper will focus on.

1.3. Adding a Flexible Form for Heteroscedasticity (HG)

To complete our econometric specification, we must use a separate instrument which will, in accordance with the strategy mentioned above, allow for a representation of the variance of the residuals. To do this, we first write a general heteroscedasticity function and then give it a useful form. The function is :

$$u_t = [f\ (Z_{1t},\ldots,Z_{Mt})]^{1/2}\ v_t \quad , \tag{3b}$$

where

$$f\ (Z_{1t},\ldots,Z_{Mt}) = \{\lambda_u\ [\delta_o + \sum_{m=1}^{M} \delta_m\ Z_{mt}^{(\lambda_{zm})}\ ] + 1\}^{1/\lambda_u} \tag{7}$$

and the Z set of explanatory variables may have common elements with the X set used in (2b). In this expression, direct BCT are applied to the Z variables which explain error variance and an inverse power transformation is applied to the expression in the square parentheses. The constant δ_o is necessary to preserve the invariance of the expression to units of measurement of the Z's [Schlesselman (1971)]. Equation (7) contains as special cases most schemes of traditional interest.

The first family of special cases is based on setting $\lambda_u = 0$. This yields

$$E(u_t^2) = \omega_{tt} = \psi^2 \exp\{\delta_o + \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\}$$

$$= \phi^2 \exp\{\sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\} \qquad (8)$$

where

$$\phi^2 = \psi^2 \exp\{\delta_o\}.$$

Furthermore, if each λ_{zm} is also set equal to zero, the last expression of equation (8) reduces to the multivariate multiplicative model found in Dagum and Dagum (1974):

$$\omega_{tt} = \phi^2 \prod_{m=1}^{M} Z_{mt}^{\delta_m} \quad . \qquad (9)$$

Equation (6) finally reduces to the classical multivariate form:

$$\omega_{tt} = \phi^2 \prod_{m=1}^{M} Z_{mt}^2 \qquad (10)$$

when each δ_m is set equal to 2. Park's (1966) specification is obtained from (8) by setting all δ_m, but one, equal to zero and by setting the λ_{zm} of the retained variable equal to zero. Imposing the further constraint that the retained δ_m equals 2 yields the classical univariate form $\phi^2 Z_{mt}^2$. Egy and Lahiri (1979) have recently used Park's specification with an equation of form (2b). Seaks and Layson (1983) present a slightly different formulation from that of Park : they suggest the form $\phi^2 Z_{mt}^{(\lambda_{zm})}$ and provide a numerical illustration which sets a unique λ_{zm} equal to a unique λ_{xk} in (2b).

Another family of special cases is obtained by setting $\lambda_u = 1$. This results in:

$$\omega_{tt} = \tilde{\delta}_o + \sum_{m=1}^{M} \tilde{\delta}_m Z_{mt}^{(\lambda_{zm})} \qquad (11)$$

where $\tilde{\delta}_o = \psi^2 (\delta_o + 1)$, $\tilde{\delta}_m = \psi^2 \delta_m$. This family includes the linear form (each $\lambda_{zm} = 1$), such as that found in Glejser (1969), and many nonlinear forms (e.g. square root, if each $\lambda_{zm} = 0.5$ and second degree, if each $\lambda_{zm} = 2$) as special cases. Buse (1979) recently formulated a different definition of generalized heteroscedasticity which also includes the linear form as a special case.

The subclass defined by (11) may be less useful than the subclass defined by (8), because the former can produce negative values for the ω_{tt}'s. We shall therefore use the latter in the rest of this paper as the useful form of (3b).

1.4. The Practical Specification of the Expanded Model

Grouping (2b), (8) and (4) yields the practical specification of the model proposed in Gaudry and Dagenais (1978, 1979) :

$$\left.\begin{aligned} y_t^{(\lambda_y)} &= \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t && (12\text{-A}) \\ u_t &= (\exp\{\sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\})^{1/2} v_t && (12\text{-B}) \\ v_t &= \sum_{\ell=1}^{r} \rho_\ell v_{t-\ell} + w_t && (12\text{-C}) \end{aligned}\right\} \qquad (12)$$

where w_t has zero mean, constant variance σ^2 and is independent among observations.

This specification potentially contains a very large number of parameters. In our present applications to three urban travel demand models with 181 monthly observations from December 1956 until December 1971, we will report full results for specifications simultaneously containing as much as 27 β_k, a λ_y and 2 λ_{xk} on the fixed part (12-A), with 2 δ_m and 2 λ_{zm} on the heteroscedasticity component (12-B) and 12 ρ_ℓ on the autocorrelation component (12-C) of the stochastic part. We will also partially report on our overall numerical experience to date with the L-1.1 algorithm used to estimate the parameters of the likelihood function and other useful statistics for model (12).

Our present applications will try to answer 3 kinds of questions : the first relate specifically to the usefulness of flexible nonlinear dimensions which depart from the basic classical multivariate least-squares model; the second examines the interdependence among these dimensions and the third concentrates on elasticities which involve all parameters of the system.

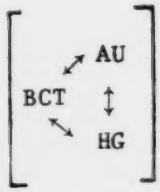
Flexible format questions :

- On the systematic part, is the *Box-Cox transformation* more useful than predetermined forms ? (BCT)
- Is a *heteroscedasticity* model of a *general* form more useful than a more restrictive form ? (HG)
- Is a full *autocorrelation* structure more useful than a sparse form ? (AU)

} (FF)

Interaction questions :

- Does the estimated form of the BCT, HG and AU component estimates each depend on the constraints imposed on the flexible form of the other components (including their suppression) ?



- More formally, are the BCT, HG and AU estimated flexible form structures each conditional on the simultaneous estimation of flexible forms in the other dimensions ? For instance, does the functional form estimated for the fixed part depend on the simultaneous estimation of autocorrelation ? and the reverse ?

(e.g.)
(BCT ← AU)
(BCT → AU)

} (INT)

Elasticity questions :

- Do the signs and values of the elasticities of the dependent variable with respect to the explanatory variables depend on the flexible format(s) used ?

$(\eta_{y,X_k} \leftarrow FF)$ } (E)

Before answering these questions with greater scope than the available literature, which is only focused on (BCT), we must formulate and justify the likelihood function and outline the principal features of L-1.1, the algorithm developed especially for the expanded model (12).

2. THE LIKELIHOOD FUNCTION

Assuming that the w_t's in (12) are normally distributed and neglecting the first r observations[2], the log-likelihood of the remaining N-r observations on the y_t's may be expressed as follows :

$$\mathcal{L} = \sum_{t=1+r}^{N} \mathcal{L}_t = -\frac{N-r}{2} \ln (2\pi\,\sigma^2) - \frac{1}{2} \sum_{t=1+r}^{N} \ln f(Z_t)$$

$$- \frac{1}{2\sigma^2} \sum_{t=1+r}^{N} \left[\frac{u_t}{\sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \frac{u_{t-\ell}}{\sqrt{f(Z_{t-\ell})}} \right]^2 + (\lambda_y - 1) \sum_{t=1+r}^{N} \ln y_t \qquad (13)$$

where

$$f(Z_t) = \exp \left\{ \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})} \right\} \quad . \qquad (14)$$

Note that the last term of the RHS of (13) is the Jacobian of the transformation from the density of the $y_t^{(\lambda_y)}$'s to that of the y_t's and that the expression in the square brackets corresponds to w_t.

2.1. A Two-limit Tobit Approach

Since the BCT on y_t is defined only for $y_t > 0$, so that equation (12-A) can hold only for values of w_t which are consistent with this constraint, it

---

2. Note that, when the number of observations is small, it might be advantageous to recuperate the first r observations, as pointed out by Beach and MacKinnon (1978). Recuperating the first r observations does not raise any theoretical problem, but the expressions of the log-likelihood function and of its first derivatives then become more messy: we have therefore not considered this alternative.

might be thought that equation (13) is not strictly valid. We maintain, however, that *it is indeed the appropriate likelihood function*. Our argument is based on the fact that in very many econometric applications, it is reasonable to consider y_t as *limited* both *downwards* and *upwards*, even if the sample contains no limit observations.

Several statistical interpretations of the BCT regression model are possible. For example, Bickel and Doksum (1981) assume $-\infty < y_t < \infty$, $\lambda > 0$ and use an extended definition of the power family, namely

$$y^{(\lambda)} = [|y|^{\lambda} \text{ sgn } (y) - 1]/\lambda \quad .$$

Bickel and Doksum note that "if in fact, the y's [in the sample] are positive, the likelihood function is as in Box and Cox". Adopting the Bickel-Doksum approach would therefore establish that equation (13) is indeed the appropriate likelihood function. Although Carroll and Ruppert (1981) are less explicit about their own assumptions, their use of the Bickel and Doksum asymptotic results suggests that they adopt essentially the same approach. However, for many econometric applications such as demand analysis, the assumption that y can take negative values is unrealistic; and for this reason, the Bickell-Doksum approach does not seem appropriate for our purpose.

Others, such as Spitzer (1976), Olsen (1978), Poirier (1978), Amemiya and Powell (1980), postulate that y_t cannot be negative and that the Box-Cox maximand is not strictly a likelihood function because it takes no account of the fact that y_t is truncated. This viewpoint is true if one (implicitly)

requires that the distribution function of $y_t^{(\lambda)}$ be everywhere continuous for all possible values of the explanatory variables and consequently contain no mass point. We shall not adopt this approach because we believe that, for some conceivable values of the explanatory variables, demand could very well take the value zero (or some value arbitrarily close to zero) with non negligible probability. In such cases, it is perfectly possible, for instance, for the distribution of $y_t^{(\lambda)}$ to have a mass point at $-1/\lambda$ when $\lambda > 0$, for particular values of the explanatory variables.

In a large number of econometric applications, such as demand analysis, it is indeed very reasonable to assume that the dependent variable has both a lower and an upper limit, even if these points are almost never actually observed. In one of the equations used in this paper, the dependent variable is the quantity of bus and subway trips per month by schoolchildren in Montreal. If the transit fare were increased from its present value of 25 cents to 40 $ or 60 $ —which is considerably higher than taxi fares within the urban area—, it is reasonable to assert that the demand for bus travel by schoolchildren in Montreal would be reduced to zero. (In practice, it is convenient, as we shall see later, to use as lower limit ε, an arbitrarily small positive number, rather than zero because the BCT can only be effected on strictly positive variables since it includes the log transformation : $\lambda = 0$). On the other hand, if the transit fare were set at zero or even at a negative value, —which would imply that the Montreal Transit Authority pay children to use transit— the demand would surely increase considerably but could not increase strictly to infinity. Given the number of available transit vehicles and the number of schoolchildren in Montreal, the monthly travel demand by schoolchildren could surely never be larger than some theoretical finite upper limit, say L.

In fact, one could argue that most standard regression analyses made in economics and in many other fields pertain to doubly limited dependent variables. This fact is usually ignored because the samples contain no limit observations and the analyses and predictions usually concern values of y_t or $E(y_t)$ which are far from both the upper and the lower limits.

The proper statistical model to use, in our case, corresponds therefore to the two-limit Tobit model [Rosett and Nelson (1975)] :

$$y_t^{(\lambda_y)} = \varepsilon^{(\lambda_y)} \quad \text{if} \quad \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t \leq \varepsilon^{(\lambda_y)}, \tag{15}$$

$$y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t \quad \text{if} \quad \varepsilon^{(\lambda_y)} < \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t < L^{(\lambda_y)}, \tag{16}$$

$$y_t^{(\lambda_y)} = L^{(\lambda_y)} \quad \text{if} \quad \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t \geq L^{(\lambda_y)}. \tag{17}$$

2.2. In the Absence of Limit Observations

If the actual sample contains no limit observations, the *exact likelihood function* reduces to that suggested by Box and Cox for regression models with homoscedastic and serially uncorrelated error terms and *to our equation (13)* for models involving heteroscedasticity and serial correlation. This can be readily inferred by considering, for example, equation (10) of the Rosett-Nelson paper and setting the number of limit observations in the sample equal to zero.

Under the assumption that the true model of $y_t^{(\lambda)}$ is a two-limit Tobit model, and given that the actual sample available contains no upper or lower limit observations, *the parameter estimators that maximize equation (13) are the true maximum likelihood estimators*[3]. What then, are the asymptotic properties of these estimators ? Different possible cases must be considered here. In line with the remarks made by Draper and Cox (1969) about the original BCT model, if it can be safely assumed that the parameters of the model, the initial conditions and the empirical distribution of the exogenous variables are such that the probability that $y_t^{(\lambda)}$ be equal to its higher or lower limit is always negligible, then it is conjectured that "the general large-sample theorems about the sampling distributions of maximum likelihood estimates", as Box and Cox (1964, p. 219) put it, apply to $\mathcal{L}_{max}$ of equation (13)[4]. Note that, as also suggested by Draper and Cox and reiterated by Olsen (1978), "the reasonableness of the [above] assumptions ... can be tested by an analysis of the residuals".

---

[3]. Strictly speaking, they are conditional maximum likelihood estimators, conditional on the values of the first r observations. Similarly, when we speak of the probability that $y_t^{(\lambda)}$ be equal to its upper or lower limit, or when we discuss the calculation of $E(y_t)$, these are all *conditional* on the past values of y.

[4]. This point has also been recognized by other authors in the literature, for the original Box-Cox model. For example, Zellner (1971, p. 163, fn. 2) notes that the range of $y_t^{(\lambda)}$ is not $-\infty$ to $+\infty$ "but just a subinterval of this range". He then points out that if the probability that $y_t^{(\lambda)}$ "will fall in the excluded interval is small, the normality assumption will not be vitally affected". In the same line of thought, it is readily seen in Amemiya and Powell (1980, p. 6, equation 2.10) that, if it is assumed that $y_t^{(\lambda)}$ has a truncated normal distribution and if the surface under the normal curve for the admissible region is close to one, the asymptotic bias of the "Box-Cox MLE" of λ vanishes; in their notation, if

$$\Phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right) \to 1 \text{ for every } t \text{ , then } -\frac{1}{\sigma_o\lambda_o^2}\sum \frac{\phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)}{\Phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)} \text{ will} \to 0 \ .$$

Clearly, a rigorous analytical derivation of the asymptotic properties of the proposed estimators of the model discussed in the present paper would require further investigation. An expensive alternative approach, which could generate information on the finite sample properties as well as on the asymptotic properties of the estimators, would be to resort to Monte Carlo experiments.

For that purpose, the probability that $y_t^{(\lambda)}$ be equal to its lower limit or to its upper limit should be estimated for each observation in the sample, as illustrated in our applications (see Section 4).

Even if it cannot be safely assumed that the probability that $y_t^{(\lambda)}$ reaches its lower or upper limit is always negligible, this probability may be sufficiently small that the probability of observing samples without limit observations remains very high. Then, in such cases, our maximum likelihood estimators remain presumably consistent and asymptotically normally distributed, given that Amemiya has shown that the maximum likelihood estimates of the Tobit model possesses such asymptotic properties[5].

Finally, if the probability that $y_t^{(\lambda)}$ be equal to its upper or lower limit is so high —given the values of the explanatory variables observed in the actual sample— as to make it unlikely that any sample of the size considered contains no limit observation, then one might clearly suspect that the suggested model is unreasonable. Note that, as pointed out above, the problem we are discussing at present is not specific to the BCT model, since it could occur with any linear regression model with a non negative dependent variable. Although this fact is most often neglected, it would be just as justified and imperative to examine the probability that the dependent variable be zero in every linear regression model with non negative dependent variable as it is for the BCT model.

---

[5]. In Amemiya's (1973) paper, the proofs that are developed seem perfectly compatible with the fact that a given *finite* observable sample may contain zero limit observations. The difficulty of extending Amemiya's proof to the model discussed in the present paper comes mainly from the assumption of serial correlation. The introduction of serial correlation in limited dependent variable models complicates greatly the general expression of the likelihood function, since every run of m limit observations leads to the introduction of an m-dimensional integral of the multivariate normal density function [Dagenais (1982)]. However, Robinson's approach (1982) could possibly be used in this case.

2.3. The Dependent Variable : its Expected Value and Probability of Being at the Limits

How then should one compute the probability that the sample contain limit observations ? We will eventually have to calculate $E(y_t)$ in order to obtain the elasticity of the dependent variable with respect to the exogenous variables and an appropriate analog of the multiple correlation coefficient R^2. The presently outlined procedure will serve all of these ends.

For each observation, the probability that y_t be equal to its lower and higher limits is readily obtained as the first and third integrals of the following expression which defines $E(y_t)$ by generalizing and extending Tobin's (1958) formula to the case of a doubly limited dependent variable $\varepsilon \leq y_t \leq L$:

$$E(y_t) = \varepsilon \int_{-\infty}^{\tilde{w}_t} f(w;\sigma^2)\, dw + \int_{\tilde{w}_t}^{w_t^*} y(w)_t\, f(w;\sigma^2)\, dw + L \int_{w_t^*}^{\infty} f(w;\sigma^2)\, dw \qquad (18)$$

where

$$\tilde{w}_t = \frac{\varepsilon^{\lambda_y} - 1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell\, y^*_{t-\ell} - \sum_{k=1}^{K} \beta_k\, X^{**}_{kt} \quad ,$$

$$w^*_t = \frac{L^{\lambda_y} - 1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell\, y^*_{t-\ell} - \sum_{k=1}^{K} \beta_k\, X^{**}_{kt} \quad ,$$

$$Z_t = (Z_{1t}, \ldots, Z_{Mt}) \quad ,$$

$$y^{*}_{t-\ell} = y^{(\lambda_y)}_{t-\ell} / \sqrt{f(Z_{t-\ell})} \quad ,$$

$$X^{**}_{kt} = X^{(\lambda_{xk})}_{kt} / \sqrt{f(Z_t)} - \sum_{\ell=1}^{r} \rho_\ell \, X^{(\lambda_{xk})}_{k,t-\ell} / \sqrt{f(Z_{t-\ell})} \quad ,$$

$$f(w;\sigma^2) = (2\pi)^{-1/2} \, \sigma^{-1} \exp\{-w^2/2\sigma^2\} \quad ,$$

and

$$y(w_t) = [1 + \lambda_y \sqrt{f(Z_t)} \, (\sum_{\ell=1} \rho_\ell \, y^{*}_{t-\ell} + \Sigma \, \beta_k \, X^{**}_{kt} + w_t)]^{1/\lambda_y} \quad .$$

2.3.1. Limiting behavior of the expected value

Let us now successively examine, over the whole domain of λ_y, how the three terms on the RHS of (18) behave as ε gets closer to zero and as L gets larger. These terms will be referred to, below, as expressions [1], [2] and [3] respectively from left to right.

When $\lambda_y > 0$, expression [1] $\to 0$ as $\varepsilon \to 0$. Similarly, expression [3] $\to 0$, as $L \to \infty$ since, for sufficiently large L, we necessarily have

$$L \int_{w^{*}_t}^{\infty} f(w;\sigma^2) \, dw \leq L \int_{w^{*}_t}^{\infty} w \, f(w;\sigma^2) \, dw \qquad (19)$$

and

$$L \int_{w^{*}_t}^{\infty} w \, f(w;\sigma^2) \, dw \to 0 \text{ as } L \to \infty \quad .$$

Therefore, when both $\varepsilon \to 0$ and $L \to \infty$, we have:

$$E(y_t) = \int_{w_t^{**}}^{\infty} y(w)_t \, f(w;\sigma^2) \, dw \qquad (20)$$

where

$$w_t^{**} = -\frac{1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \, y^*_{t-\ell} - \sum_{k=1}^{K} \beta_k \, X^{**}_{kt} \quad .$$

The value of $E(y_t)$ can then be computed by numerical methods.

When $\lambda_y = 0$, expressions [1] and [3] both tend toward zero as $\varepsilon \to 0$ and $L \to \infty$. In addition term [2] can be rearranged to yield:

$$E(y_t) = \int_{-\infty}^{\infty} \bar{y}(w)_t \, f(w;\sigma^2) \, dw \qquad (21)$$

where

$$\bar{y}(w)_t = \exp\{\sqrt{f(Z_t)} \, (\sum_{\ell=1}^{r} \rho_\ell \, y^*_{t-\ell} + \sum_{k=1}^{K} \beta_k \, X^{**}_{kt} + w_t)\} \quad .$$

Note therefore that, when $\lambda \geq 0$, the fact that y_t may have an upper limit is of no consequence in practice for the computation of $E(y_t)$, provided one assumes that this upper bound is sufficiently high. Indeed, for a sufficiently large value of L, the probability associated with the mass point at L tends toward zero for any finite values of X^{**}_{kt} $(k = 1,\ldots,K)$, $y^*_{t-\ell}$ $(\ell = 1,\ldots,r)$ and the elements of Z_t; furthermore, the value of w_t^* in [2] can be replaced by ∞ without altering

the value of the integral. Therefore, when $\lambda_y \geq 0$, the upper bound L does not even have to be specified explicitly, provided it is assumed to be sufficiently large[6].

The case is different however for $\lambda_y < 0$. This case cannot be treated as the two previous cases since, if we assume that L get arbitrarily large and $\to \infty$, $E(y_t)$ does not exist. In this case, the value of L must be specified explicitly if one wants to compute $E(y_t)$. Therefore, when $\lambda < 0$, as $\varepsilon \to 0$, term [1] of (18) $\to 0$ and $E(y_t)$ becomes:

$$E(y_t) = \int_{-\infty}^{w_t^*} y(w)_t \, f(w;\sigma^2) \, dw + L \int_{w_t^*}^{\infty} f(w;\sigma^2) \, dw \quad . \qquad (22)$$

Note that, for values of Z_t, X_t^{**} and y_t^* such that $E(y_t)$ is much smaller than L, the probability that y_t attain its upper limit L will be negligible and the value of $E(y_t)$ will be quite insensitive to the precise value retained for L. In samples where no upper limit values of y_t have been observed, one may infer that all the Z_t, X_t^{**} and y_t^* values contained in the sample are effectively such that $E(y_t)$ is much smaller than L for every t. Hence, it should be possible to compute $E(y_t)$ for the observations

---

[6]. In certain cases, one may not be ready to make such an assumption and might prefer to specify the value of L explicitly. In such case, the expression for $E(y_t)$ would become:

$$\text{for } \lambda_y > 0 \; : \; E(y_t) = \int_{w_t^{**}}^{w_t^*} y(w)_t \, f(w;\sigma^2) \, dw + L \int_{w_t^*}^{\infty} f(w;\sigma^2) \, dw$$

and

$$\text{for } \lambda_y = 0 \; : \; E(y_t) = \int_{-\infty}^{\bar{w}_t} \bar{y}(w)_t \, f(w;\sigma^2) \, dw + L \int_{\bar{w}_t}^{\infty} f(w;\sigma^2) \, dw$$

where

$$\bar{w}_t = \frac{\ln L}{\sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \, y_{t-\ell}^* - \sum_{\ell=1}^{K} \beta_k \, X_{kt}^{**} \quad .$$

contained in the sample to the desired level of accuracy without requiring that the value of L be known with great precision : it should be sufficient to specify L only very approximately[7].

2.3.2. Probability that the sample contains limit observations

Having established that it is possible to obtain a value for $E(y_t)$ over the whole domain of λ_y and for all meaningful values of the lower and upper limits, how is one to use the probability of being at the lower limit obtainable as the value of the integral in [1] and the probability of being at the higher limit obtained as the value of the integral in [3] ?

These probabilities can in principle be examined for each observation t but a summary measure is more convenient. Moreover, when λ_y is positive, the fact that y_t may have an upper limit is of no practical consequence since the probability associated with the mass point tends to zero for sufficiently large values of L : our concern in this case should be with the probability that y_t be equal to its lower limit zero, unless our sample pertained to a phenomenon, such as the number of work hours per day, where the upper limit is naturally low. Similarly, when λ_y is negative, the

---

7. Because of these considerations, in our program [Liem et al. (1983)], the user is asked to specify the value of L. Otherwise, if the user *does not* specify L and the maximum likelihood estimate of λ_y turns out to be *negative*, the program generates, by default, its own approximation of L and uses it to compute $E(y_t)$ from equation (22). When the L value is thus generated by the program, it is automatically printed in the output, so that the user may be made aware of the approximation actually used. This procedure is further discussed in Liem et al. (1983).

probability associated with the lower limit mass point tends to zero for sufficiently small values of ε : we should in this case be concerned with the probability that y_t be equal to its upper limit unless our sample contained a significant positive lower limit. When $\lambda_y = 0$, probabilities associated with both upper and lower limits tend to zero for sufficiently small values of ε and sufficiently large values of L.

In consequence, we will compute over the whole sample the following statistics :

$$P_1 = \prod_{t=1+r}^{N} \Pr(y_t > \varepsilon) = \prod_{t=1+r}^{N} \left(1 - \int_{-\infty}^{w_t(\varepsilon)} f(w;\sigma^2)\, dw\right) , \quad \text{for } \lambda_y > 0 \qquad (23\text{-A})$$

and

$$P_2 = \prod_{t=1+r}^{N} \Pr(y_t < L) = \prod_{t=1+r}^{N} \left(1 - \int_{w_t(L)}^{\infty} f(w;\sigma^2)\, dw\right) , \quad \text{for } \lambda_y < 0 . \qquad (23\text{-B})$$

These statistics are clearly very sensitive to the value of the highest probability of being at the limit found in the sample and are therefore conservative indicators of the appropriateness of using the likelihood function with a given sample.

3. THE L-1.1 ALGORITHM

In order to maximize $\mathcal{L}$ in equation (13) with respect to the vector of unknown parameters π, where $\pi = (\lambda_y, \lambda_{x1},\ldots,\lambda_{xK}, \lambda_{z1},\ldots,\lambda_{zM}, \delta_1,\ldots,\delta_M, \rho_1,\ldots,\rho_r, \beta_1,\ldots,\beta_K, \sigma^2)'$, the function is first concentrated over σ^2 and the β's. Then the concentrated likelihood function $\bar{\mathcal{L}}(\bar{\pi})$, where $\bar{\pi} = (\lambda_y, \lambda_{x1},\ldots,\lambda_{xK}, \lambda_{z1},\ldots,\lambda_{zM}, \delta_1,\ldots,\delta_M, \rho_1,\ldots,\rho_r)'$, is maximized with respect to the elements of $\bar{\pi}$, using the Davidon-Fletcher-Powell (DFP) algorithm [see Fletcher and Powell (1963)].

3.1. The Concentrated Likelihood

To obtain $\bar{\mathcal{L}}$, concentration over the β's is obtained by writing $\frac{\partial \mathcal{L}}{\partial \beta} = 0$, where $\beta = (\beta_1,\ldots,\beta_K)'$, and solving for β to get:

$$\bar{\beta} = (X^{**\prime}X^{**})^{-1} X^{**\prime} y^{**} \tag{24}$$

where

$$X^{**} = (X^{**\prime}_{r+1},\ldots,X^{**\prime}_{N}) \quad ,$$

$$X^{**}_{t} = (X^{**}_{1t},\ldots,X^{**}_{Kt}) \quad , \qquad (t = r+1,\ldots,N)$$

$$y^{**} = (y^{**}_{r+1},\ldots,y^{**}_{N})' \quad ,$$

and

$$y^{**}_{t} = y_t^{(\lambda_y)} / \sqrt{f(Z_t)} - \sum_{\ell=1}^{r} \rho_\ell \, y_{t-\ell}^{(\lambda_y)} / \sqrt{f(Z_{t-\ell})} \quad .$$

In a second step, concentration over σ^2 is achieved by writing $\partial \mathcal{L}/\partial \sigma^2 = 0$, solving for σ^2 and replacing β by $\bar{\beta}$. The solution is:

$$\bar{\sigma}^2 = [1/(N-r)] \sum_{t=r+1}^{N} [y^{**}_{t} - \sum_{k=1}^{K} (\bar{\beta}_k \, X^{**}_{kt})]^2 \quad . \tag{25}$$

Substitution of $\bar{\beta}$ and $\bar{\sigma}^2$ in $\mathcal{L}$ yields the concentrated log-likelihood

$$\bar{\mathcal{L}} = -[(N-r)/2]\{\ln(2\pi) + 1\} - 1/2 \sum_{t=1+r}^{N} \ln f(Z_t)$$

$$- [(N-r)/2] \ln \bar{\sigma}^2 + (\lambda_y - 1) \sum_{t=1+r}^{N} \ln y_t \quad . \tag{26}$$

3.2. Maximization and Determination of the Autoregressive Structure (AU)

In order to apply the iterative DFP gradient method, the derivatives of $\mathcal{L}$ with respect to the elements of $\bar{\pi}$ have to be computed at each iteration. The analytical expressions of these derivatives are given in Liem et al. (1983). An initial solution must also be supplied to start the iterative process. Setting the λ_x's, the λ_z's and λ_y equal to one and the ρ's as well as the δ's equal to zero appears to be a good starting point, since the model reduces to a classical linear multiple regression.

After a first estimate of β has been found by using equation (24), a first estimate of the u's is obtained from equation (1) and a Box-Jenkins (1970) type analysis applied to help determine the autoregressive (AR) structure of the error terms when the sample is sufficiently large. Although the actual AR structure of the estimated residuals could change between iterations, in practice the AR structure identified at the first iteration is retained for all subsequent iterations. Once the maximum has been attained, the AR structure of the estimated w's is examined again. We have made several numerical experiments and, in almost all cases, the estimated w's obtained at the final iteration appeared to behave as a pure white noise[8]. If the sample is small, the Box-Jenkins techniques may not be very useful for identifying the AR pattern which must consequently be determined *a priori* by the experimenter.

---

[8]. If that was not the case, it was usually easy to change the AR structure in such a way as to obtain the desired result, by trial and error. The reader should also note our Section 4.3.3 on this point.

3.3. t-Statistics : Unconditional but Dependent on Units of Measurement

To estimate the asymptotic covariance matrix of $\hat{\pi}$, the maximum likelihood estimate of all the elements of π, the formula suggested by Berndt et al. (1974) is used. This requires computing the derivative of each $\mathcal{L}_t$ defined by equation (13) with respect to each element of π evaluated at $\hat{\pi}$. The gradient so obtained can be written as :

$$\frac{\partial \mathcal{L}}{\partial \pi} = \sum_{t=1+r}^{N} \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right)_{\pi=\hat{\pi}} \tag{27}$$

where

$$\frac{\partial \mathcal{L}_t}{\partial \pi} = (\ln y_t) \frac{\partial(\lambda_y - 1)}{\partial \pi} - \frac{1}{2} \frac{\partial \ln f(Z_t)}{\partial \pi} - \frac{w_t}{\sigma^2} \frac{\partial w_t}{\partial \pi}$$

and w_t denotes the contents of the square bracket in (13).

The covariance matrix is expressed as:

$$\text{cov} = \left[\sum_{t=1+r}^{N} \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right) \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right)' \right]^{-1}_{\pi=\hat{\pi}} . \tag{28}$$

The asymptotic t-statistics are obtained in usual fashion by computing the ratio of each parameter to the square root of the appropriate element from the diagonal of (28). The t-values thus obtained are unconditional : they take into account all estimated parameters of the system. One should however keep in mind the following points when examining the forthcoming tables of results which present the t-statistics :

1) Wald-type tests such as asymptotic t-tests are not generally invariant to reparametrization [Burguette, Gallant and Souza (1982)]. In particular, in models involving Box-Cox transformations, the t-statistics associated with the β parameters are not invariant to the reparametrization implied by changes in the units of measurement [Spitzer (1984];

2) conditional t-statistics for the β's, for given values of the λ's, are invariant but, in certain extreme cases, these may be considerably larger (in absolute value) than the corresponding, more appropriate, unconditional statistics [Spitzer (1982), Blackley, Follain and Ondrich (1984)];

3) if the units of measurement of y and of the X's are standardized, so that all experimenters utilizing the same model and the same sample adopt the same units of measurement, this will clearly eliminate the possibility that different experimenters use different units of measurement and get different results for the t-tests. For example, each variable can be divided by its sample arithmetic or geometric [Spitzer (1984)] mean. When the arithmetic mean is used, the estimates of the β's in the standardized model coincide with the estimates of the "intuitive elasticities" evaluated at the means (to be defined later). The adoption of such standardization procedures, however, does not resolve the fundamental problem arising from the fact that, in BCT models, asymptotic t-test are not invariant with respect to changes in the units of measurement of the variables; this approach simply avoids the question by imposing an arbitrary norm[9].

---

[9]. The fact that the norm defined for y (namely its sample arithmetic mean $\bar{y}$ or its sample geometric mean $\dot{y}$) is itself a stochastic variable also raises statistical complications which have not been considered, for example, by Spitzer (1984).

Even if presenting t-statistics for the β's conditional on the values of the λ's corresponding to their maximum likelihood estimates, or standardizing the y's and X's by dividing them by their sample arithmetic means, offer no adequate solution to the problem of lack of invariance of the t-statistics[10], we shall present in the Appendix such results for one of the models discussed below. In that particular case, the conditional tests on the β_k differ markedly from the unconditional tests originally obtained, but the test on the variables scaled by their arithmetic means are, in general, not very different however from those presented in Table 1.

The solution to this problem may well lie in the development of confidence regions for the "strict elasticity" (to be defined later) of the dependent variable with respect to the X_k and the Z_m in (12). As noted below, elasticities, rather than coefficients as such, are of interest in a model. We conjecture that t-tests on properly defined "strict elasticities" might be invariant to units of measurement.

3.4. Analogs of the Multiple Correlation Coefficient R^2

In linear regression analysis, a useful statistic is the multiple correlation coefficient R^2 which relates the $\hat{y}$ values "predicted" by the model to the observed values :

$$R^2 = 1 - \sum_{t=1+r}^{N} (y_t - \hat{y}_t)^2 / \sum_{t=1+r}^{N} (y_t - \bar{y})^2 \quad . \tag{29}$$

---

[10]. Note however that likelihood ratio tests and Rao-type tests [Rao (1948)] are invariant with respect to changes in the units of measurement. Performing such tests for each individual β parameter would however be much too costly, in our case.

It would be convenient to have a similar measure for nonlinear models. In our case, for instance, one might wish to extract $\hat{y}$ by "unrolling" the BCT on y; from 12-A, such an inversion would yield :

$$\hat{y}_t = [\hat{\lambda}_y \left(\sum_{k=1}^{K} \hat{\beta}_k X_{kt}^{(\hat{\lambda}_{xk})} \right) + 1]^{1/\hat{\lambda}_y} \quad , \tag{30}$$

where it is clear that $\hat{y}_t$ might not be a real number because the contents of the square brackets need not be positive[11]. We have therefore defined the following two measures of adjustment :

$$\text{Pseudo-(E)-}R^2 = 1 - \sum_{t=1+r}^{N} [y_t - E(y_t)]^2 \Big/ \sum_{t=1+r}^{N} (y_t - \bar{y})^2 \tag{31}$$

and

$$\text{Pseudo-(L)-}R^2 = 1 - \exp\{\frac{2}{N} (\mathcal{L}_o - \mathcal{L})\} \quad , \tag{32}$$

where

$\mathcal{L}$ is the value of the log-likelihood of the model considered and $\mathcal{L}_o$ is the value of the log-likelihood under the hypothesis that the model is linear, without autocorrelation and heteroscedasticity, and that all β_k are equal to zero except the constant.

Although both (31) and (32) coincide with (29) in the classical linear regression case with spherical residuals, in the present case, only (32), which increases monotonically with $\mathcal{L}$, is necessarily comprised between 0 and 1. Our program computes both but we will only report (32) in our tables of results.

---

[11]. It is unclear how this unrolling is performed in UBC SHAZAM [White and Ho, (1982), p. 58].

4. RESULTS : FUNCTIONAL FORM, AUTOCORRELATION AND HETEROSCEDASTICITY

In this section, we will report mostly on statistical results used to answer the first two kinds of questions outlined in Section 4.1. : (FF) questions concerning the usefulness of each of the three nonlinear dimensions in which model (12) expands the classical linear regression and (INT) questions concerning the interaction among these dimensions. To simplify the presentation, we will refer to each of the demand equations as follows :

Equation A = transit demand by adults,
Equation S = transit demand by schoolchildren,
Equation G = gasoline demand.

} (Code 1)

4.1. A Table of Likelihood and BCT Values

Our principal source of information to answer the questions of interest will be Table 1 which contains 120 model specification results (102 if repetitions are excluded). Each result consists of the values of the log-likelihood and of the BCT associated with the systematic part (12-A) of the specification considered. All 120 experiments have row and column identification : series numbers denote the (double-line) rows and the experiment numbers refer to columns. For instance, comparisons of nested cases can be made as follows :

(S-3, 1.1) vs (S-3, 1.3)
and
(S-3, 1.3) vs (S-4, 1.3) .

} (Code 2)

TABLE 1

LOG LIKELIHOOD VALUES (L) AND BOX-COX LAMBDA PARAMETERS (BCT) OBTAINED WITH THE L-1.1 ALGORITHM ON THE DEMTEC MODEL OF MONTHLY DEMAND FOR TRANSIT (ADULTS, SCHOOLCHILDREN) AND ON A MODEL OF MONTHLY DEMAND FOR GASOLINE IN MONTREAL (December 1956-December 1971)

Column #		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Class			LOG+				LIN+					BCT(1)+				BCT(3)+
Sub-class		---	AU	HG_1	$AU+HG_1$	---	AU	HG_1	$AU+HG_1$	---	AU	HG_1	$AU+HG_1$	$AU+HG_2$	$AU+HG_{1+2}$	$AU+H_{1+2}$
Experiment #		0.0	0.1	0.2	0.3	1.0	1.1	1.2	1.3	2.0	2.1	2.2	2.3	2.3.2	2.4	3.4
ADULTS																
A-1	BCT:	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	.75*	.92*	.75*	.91*			
(r=1,12)	L:	258.72	270.84	258.89	273.13	271.29	308.45[D]	271.41	309.44	272.56	308.73	272.60	309.75			
A-2	BCT:	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	.75*	.92*	.75*	.92*			
(r=1.12)	L:	256.81	268.87	257.01	271.26	269.30	306.42	269.46	307.47	270.53	306.64	270.58	307.73			
A-3	BCT:	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	.75*	1.10*	.76*	1.12*			
(r=1,12,13)	L:	256.81	268.91	257.01	271.27	269.30	310.92	269.46	313.11	270.53	311.49	270.59	313.83			
SCHOOLCHILDREN																
S-1	BCT:	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	.51**	.72*	.48**	1.00*	.75*	.64**	.42**1.66* 1.08*
(r=1,3,4,12)	L:	-3.95	17.55	9.18	24.58	-2.69[1]	27.02[D]	9.57	37.48	1.39[2]	28.20[3]	13.92[4]	37.48[6]	29.00[7]	53.86[8]	56.79[9]
S-2	BCT:	"	0.00	"	0.00	"	1.00	"	1.00	"	.84*	"	1.13*	.75*	.63**	(global maximum)
(r=1,3,7,9,12)	L:	"	22.87	"	29.10	"	36.37	"	48.07	"	36.69	"	48.25	36.94	53.31	(not found)
S-3	BCT:	"	0.00	"	0.00	"	1.00	"	1.00	"	.84*	"	1.11*	.79*	.76*	.83* 3.04* .92*
(r=1,3,4,7,9,12)	L:	"	26.68	"	34.52	"	41.52	"	53.40	"	41.83	"	53.56	42.00	57.86	60.44
S-4	BCT:	"	0.00	"	0.00	"	1.00	"	1.00	"	.81*	"	1.05*	.80*	.46**	.86* 4.18**1.05*
(r=1,...,12)	L:	"	30.88	"	38.70	"	45.70	"	56.96	"	46.14	"	56.99	46.17	64.86	68.07
GASOLINE																
G-1	BCT:	0.00	0.00	0.00	0.00	1.00	1.00[D]	1.00	1.00	.86*	.01*	.99*	.04*			
(r=1,12)	L:	10.54	48.97	13.30	52.99	28.00	43.94	33.64	48.79	28.55	48.98	33.65	53.01			

LIN=linear; LOG=logarithmic; BCT(p)=Box-Cox with p lambda parameters; AU=autocorrelation structure r; H-=heteroscedasticity of general (G_v), or classical (C_v), nature with variable # v. For equation groups A (=Adults), S (=Schoolchildren), and G (=Gasoline), v=1 denotes SHONF, TEND and TRCUSP, respectively, and v=2 denotes SHONF. The symbols (*), (*), or (**) denote a lambda estimate with a t-statistic higher than 2 in absolute value when the test is made, respectively, with 0, or 1, as maintained hypotheses; (**) indicates that both statistics are higher than 2. All estimates are obtained with 181 observations except for A-2 and A-3 which only use 1980. Likelihood values which have the superscript ([D]) correspond to the DEMTEC model specifications found in Gaudry (1975, 1978, 1980) and estimated with other algorithms; broken underlines (---) designate cases presented in Gaudry and Wills (1978) with estimates obtained from a less precise grid search. The numbers used as superscripts of the likelihood values, such as ([3]), are the identity numbers of the ten experiments presented in an earlier and widely circulated version of this paper (Dagenais et al., 1980): for cases 1, 2, 3, 4, 7, 8, and 9, current likelihood values shown here differ by a constant from the earlier ones except for case 6 which has been corrected; cases 5 and 10 of Dagenais et al. (1980) are not shown by want of space but are briefly discussed in the text. Complete regression results are given in Table 3 for estimates underlined twice (__) and in Table 4 for estimates underlined once (__).

Table 1 is organized in the following way :

- each of the 4 vertical sections defines a class based on the BCT specification for the systematic part : the BCT applied to all strictly positive variables is fixed at 0 in collumns 1 - 4 and at 1 in columns 5 - 8. A single common BCT ($\lambda_y = \lambda_{xk}$ for all $X_k > 0$) is estimated in columns 9 - 14 and, in column 15, two BCT are used on different groups of explanatory variables in addition to the BCT applied to the dependent variable;

- in the first vertical section, experiment 0.0 is made without taking either autocorrelation or heteroscedasticity into account; experiment 0.1 adds only autocorrelation (according to (12-C)) to reference specification 0.0 and experiment 0.2 adds only a general form of heteroscedasticity (according to (12-B)) with one Z_m variable; experiment 0.3 adds both autocorrelation and heteroscedasticity to the reference specification 0.0;

- the same pattern is used in the second vertical section and in the first 4 columns of the third section except that the reference specifications are, respectively, the linear case (1.0) and the Box-Cox specification (2.0);

- in column 13, the general heteroscedasticity form uses a Z_m variable which is not the same as that used in column 12; subscripts 1 and 2 used in column subtitles HG_1 and HG_2 indicate that fact;

- in columns 14 and 15, the two Z_m variables used in the previous two columns are used simultaneously, as denoted by the two subscripts on HG_{1+2};

- the three horizontal sections on A, S and G contain series of experiments in which a given autoregressive structure "r" is (where applicable) used; all series of results are based on 181 observations minus the highest order of autocorrelation, which yields 169 observations in most cases except for series A-2 where 168 observations are used to make it comparable to series A-3.

The cases may be nested in the following ways :

- in each of the first three vertical sections, the last column contains as nested special cases the remaining columns of the section;

- columns 9 - 12 contain as nested cases corresponding columns 1 - 4 and 5 - 8;

- columns 14 and 15 contain as nested special cases all columns to their left;

- in columns which contain autoregressive structures, series A-2 is a nested special case of series A-3, S-1 and S-2 are nested into S-3 and the latter three are special cases of S-4.

Before we specifically address the (FF) and (INT) questions, a few general comments should be made on the reasonableness of the stochastic approach and the behavior of the likelihood function.

4.2. General Comments

The following general remarks can be made about the results presented in Table 1.

4.2.1. Reasonableness of the stochastic approach

As pointed out in Section 3.1, our maximum likelihood solution is valid only if the probability that y_t be equal to its higher or lower limit is negligible. We have examined, for series A-1, A-2, S-1, S-4 and G-1, both the highest probability and the overall indicator (23-A) for all experiments which have an estimated λ —all of our estimates of λ_y are positive. For all models, the value of P_1 was equal to 1 except for (S-1, 2.3.2) where it was equal to 0.996; in that case, the highest conditional probability of being at the limit for any given observation was of the order of 10^{-3}.

This result is hardly surprising : when the $E(y_t)$ is quite different from both upper and lower limits, the finicky objections we have raised in Section 3.1. are somewhat puritanical. In many samples, one or two observations may be more likely to be at the limit than the others but that would generally not require one to reject the statistical approach : modelling judgment is required.

4.2.2. Behavior of the likelihood function

As anticipated, the likelihood function increases as constraints are relaxed and the model becomes more general. In many cases, notably in columns 12 - 15,

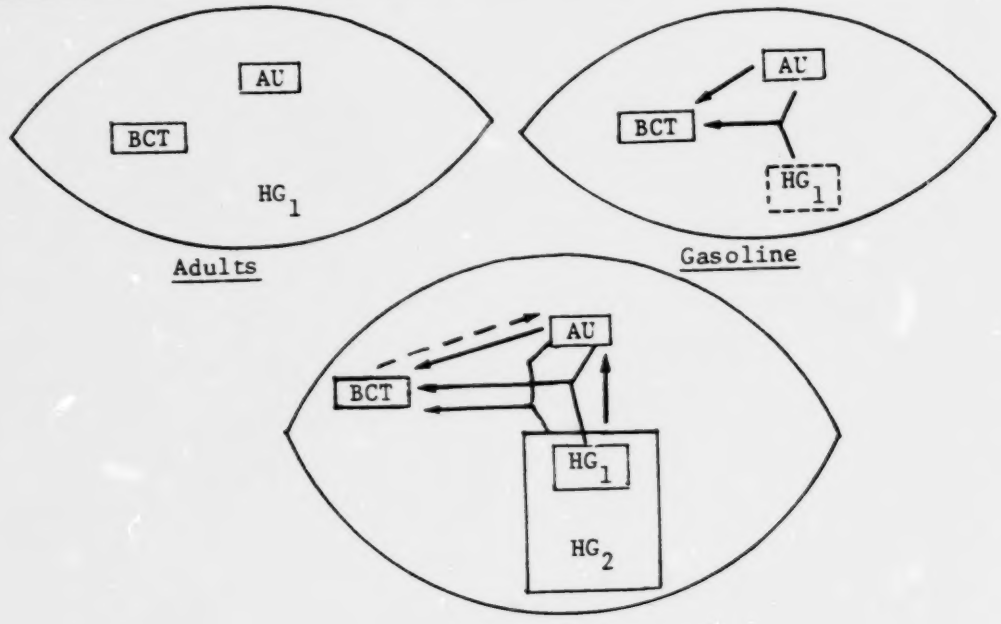
different initial solutions were tried but they always led to the same maximum; an exception is (S-2, 3.4) where we located many local minima but the global maximum, which is necessarily higher than that corresponding to (S-2, 2.4), could not be found.

4.3. Flexible Format Questions

We now want to answer the (FF) questions, considering them in turn. We are in effect asking about each of the three dimensions (BCT), (HG) and (AU) whether its existence could be established and, if so, whether its significance was unconditional or conditional upon one or both of the other two. Figure 2 summarizes our answers to the three (FF) questions without specifying which of the Table 1 cases were used to construct the figure. We will not review the cases in detail but simply give some more flesh to the information combined in Figure 2.

4.3.1. The usefulness of BCT on the systematic part

The answer to the (BCT) question is already known in the literature, as Section 1 makes clear. Table 1 uses asterisks to denote BCT estimates which are significantly different from 0, 1, or both, on the basis of their t-statistics. Column 9 shows that the logarithmic model is rejected for A, S and G equations. In the more general models of column 12, the DEMTEC equations are linear and the G equation is logarithmic. The BCT can be shown to dominate either the linear or the log form unconditionally, i.e. independently from whether one simultaneously takes into account heteroscedasticity and autocorrelation. Furthermore, no significant gain is obtained by using 3 BCT instead of 1 for S, as a comparison of column 14 and column 15 will show.



Dimensions are represented by letter codes

BCT denotes the presence of Box-Cox transformations, according to (12-A),

AU denotes the presence of autocorrelation according to (12-C),

HG_1 and HG_2 denote heteroscedasticity of general form (12-B) with variable # 1 and # 2, respectively.

Boxes denote statistical significance of the dimension in question

The absence of a box means that the dimension is not significant.

A box denotes a dimension which is unconditionally present even if the other dimensions are not taken into account.

A box made of broken lines means a dimension significant conditionally upon the presence of another dimension.

Arrows denote a significant influence on the nature or form of the indicated dimension

All arrows (and boxes) refer to phenomena the existence of which can be verified by comparing nested cases of Table 1 except the broken arrow which refers to a phenomenon which can only be established by the analysis of the autoregressive structure of experiments 2.4 and 3.4 in Table 4. (See Section 4.4.3).

DIMENSIONS AND THEIR INTERACTIONS ESTABLISHED FROM TABLE 1

FIGURE 2

In other tests with these data, we have found that we could obtain significant conditional gains with 6 BCT. Moreover, our experience with time-series data suggests that allowing at least y to be affected by a BCT, whether or not the variables on the right hand side are also affected, can be profitable, as if applying the BCT to the y's were more crucial than applying it to the X's. This may be related to the fact that λ_y affects a variable which has no β parameter (by contrast with the X's) to "compensate" in part, so that the error sum of squares tends to be immediately affected.

4.3.2. A general form for heteroscedasticity

The general form specified in (12-B) is completely new and somewhat daring because it is a powerful nonlinear transformation which, in (13), is used as a weighting factor for each observation. This means that a relatively large (or small) value of Z_m for a certain period t can transform radically the pattern of pairwise correlations (or the multicollinearity structure) among the transformed X's implicit in the square bracket of (13) and create a quasi-singular X** matrix in (24). Although we have not had this problem in the experiments of Table 1, we have seen it on other occasions. It is difficult to predict this because one does not easily have an intuitive feeling for the behavior of function (12-B).

As the summary of Figure 2 makes clear for HG_1, we could find no significant heteroscedasticity in A when we tested with a first variable; in G, HG_1 was significant, but only conditionally : (G-1, 0.2) does not dominate (G-1, 0.0),

but if a BCT is used (or a linear model is used), then the HG_1 dimension is significant. The S equation provides a rather remarkable example of the usefulness of our formulation : HG_1 is significant (column 12 vs column 10), HG_2 is not (column 13 vs column 10) but HG_1 and HG_2 used *jointly* are (column 14 vs column 12 and vs column 13)! For experiment series S-1, we have performed an experiment with "classical" heteroscedasticity (10) using the same variable as that used for HG_1 : the resulting likelihood value[12] was 17.837 units *lower* than the value shown in (S-1, 2.0), where one does not try to correct for heteroscedasticity at all! This notable decrease in likelihood suggests that the correction can be too rigid and that it is important to use an appropriate correction.

4.3.3. A multiple autocorrelation structure

- *the Box-Jenkins arts*

We have described in Sections 1.1. and 3.2. how we use Box-Jenkins methods to obtain white noise residuals by analysing both the u_t residuals (for diagnostic purposes) and the w_t residuals in (12) (to verify that the correction has been made). Using such tests, the w_t residuals of A and G experiments are approximately white noise. In S, the first series of experiments yields white noise residuals if a low degree of precision, say the 10^{-4} used in previous papers, is accepted as numerical criterion to stop the likelihood maximization process. But if a higher level of precision, say the 10^{-7} level used for all experiments of Table 1, is adopted, then the w_t residuals for the series S-1 still contain significant non random elements

---

[12]. This was case 5 in the previous version of this paper (Dagenais et al., 1980).

and the slighly different structure used for S-2 is required to obtain white noise residuals. If the Box-Jenkins "art" of interpretation is used, there is no apparent need to examine other autocorrelation structures than that of S-2.

- a conflict of tests ?

However, we performed series S-3 in order to obtain S-1 and S-2 as nested special cases and found everywhere a considerable log-likelihood gain of 4 to 5 units, when S-3 and S-2 are compared, due to the simple addition of ρ_4 ! Furthermore, the use of an additional ρ_ℓ, to "complete" the series from ρ_1 to ρ_{12}, yielded (as expected) non significant gains in all cases but (S-4, 2.4) and (S-4, 3.4) where gains of 7 units are significant at the 95% level! This evidence strongly suggests that conflicts can arise between Box-Jenkins and likelihood ratio tests, conflicts which imply that the Box-Jenkins identification procedures may not be perfectly satisfactory.

4.4. Interaction Questions

In Section 1.2, we described why the stochastic specification is important and should be considered simultaneously with the flexible form of the fixed part of a model. We emphasized the fact that transformations can influence the sphericalness of the distribution of residuals, notably the degree of heteroscedasticity, more than we emphasized the reverse. We can now afford to be more precise by answering the (INT) questions. In Figure 2, solid arrows indicate an interdependence among BCT, HG and AU dimensions only when departing from the BCT-with-spherical-residuals specification has a significant effect on the nature (the estimated parameters) of the other dimensions. For instance, the comparison of (G-1, 0.2) and (G-1, 1.2)

with their respective predetermined forms (G-1, 0.0) and (G-1, 1.0) shows that using a log-linear instead of a linear model can remove heteroscedasticity of a general form HG_1, somewhat as in case A of Figure 1, but we have no arrow from the BCT box to the HG_1 box for G in Figure 2 because *using* a BCT dimension adds heteroscedasticity only if the point of departure is a logarithmic model. The broken line arrow is an exception in the sense that the point of reference is not a linear fixed part but one which contains one BCT common to both the dependent variable and the explanatory variables; the arrow shows that permitting two λ_{xk} and allowing these to differ from λ_y modifies the AU structure.

4.4.1. The influence of AU and HG on BCT

We have found the following significant changes in BCT estimates :

- *influence of AU on BCT*

In both G and S models, adding a multiple autocorrelation structure significantly affects the estimates of the BCT on the fixed part : in the former case, one obtains a logarithmic form instead of a linear form [(G-1, 2.1) vs (G-1, 2.0)]; in the latter, a square root model with $\lambda_y = \lambda_x$ significantly different from both 0 and 1 becomes nearly linear when AU is added.

- *influence of HG on BCT*

We have not found cases in which the addition of HG significantly affected BCT estimates; but we have found cases in which the addition of HG to a model where AU is already taken into account affects BCT estimates. In G-1, such

an addition makes a linear model become logarithmic (2.3 vs 2.2). In S, the BCT estimate significantly changes when only HG_1 is added (column 12 vs column 11) and also when both HG_1 and HG_2 are added (column 14 vs column 11).

4.4.2. The influence of HG on AU

We have not found instances of significant impact of AU on the parameter values of HG, but we have found an interesting case of influence in the other direction where the addition of HG_1 and HG_2 to a model with BCT and AU changes completely the structure of autocorrelation despite the fact that the addition of either HG_1 or HG_2 has no such effect! This case, which holds for all comparisons between column 14 and 13, can only be adequately understood by looking at the signs and t-statistics of all 12 ρ's in Table 4 and noticing how the stable pattern breaks as one reaches column 2.4 of that table.

4.4.3. The influence of BCT on AU and HG

Relaxation of the constraints on BCT has a dramatic effect on the structure of AU, as a comparison of column 3.4 and column 2.4. of table 4 shows. Added to the point made in 4.4.1. above, this means that we have a bidirectional influence, or complete interaction, between BCT and AU for equation S, as indicated in Figure 2. We have not found that using many BCT influenced the nature of HG.

After answering (FF) and (INT) questions, we turn to (E) questions.

5. RESULTS : ELASTICITIES AND VALUES OF TIME

The purpose of model (12) is to obtain better estimates of the importance of explanatory factors. The most useful measure for this purpose is the elasticity of the dependent variable with respect to the explanatory variables. Showing the elasticities is generally more demanding than showing the β_k coefficients of a model in the sense that elasticities immediately betray difficulties and lack of realism of results. By contrast, β_k coefficients usually have no unambiguous meaning unless the units of measurement of the variables are specified, and they are not generally what one is interested in, anyway.

5.1. Intuitive Versus Strict Elasticity

But there are different notions of elasticity. Consider the simplified model

$$y^{(\lambda_y)} = G(X) + u, \qquad X = (X_1,\ldots,X_K) \quad , \tag{33}$$

where G(X) is a nonlinear function of X and the u's have a spherical distribution. A natural way of defining elasticities, conceived as the ratio of the percentage change in y to the percentage change in X_k, is to use the intuitive definition of point elasticity :

$$\eta_k(S) = \left[\frac{\partial y}{\partial X_k}\right]_{X=\bar{X},\, y=\bar{y}} \cdot \frac{\bar{X}_k}{\bar{y}} \quad . \tag{34}$$

Evaluation of this elasticity requires choosing a point at which to evaluate the marginal change (slope) and relating this change to base levels of X_k and y. As any point can be used to evaluate the derivative and serve as reference level, it is customary to use the same point for both purposes, usually the sample means $\bar{X} = (\bar{X}_1,\ldots,\bar{X}_K)$ and $\bar{y}$ considered as "representative" of the sample. This practice gives a result to which it is easy to relate immediately.

A first problem with the intuitive "sample" definition of elasticity (34) is that it uses the stochastic element $\bar{y}$ and is consequently a stochastic measure of sensitivity, not a fixed one[13]. Moreover, it fails to take into account the distribution function of y (through the distribution function of the error terms) in determining the relative influence of a given X_k on y. In (34), only the *sample* values of y are used to define a representative value of y : no use is made of the specified *model* to identify this value. One might then ask what parameters or properties of the distribution of y should be of interest.

A first candidate is the median. In this case, one could write the following :

$$\eta_k(M) = \left[\frac{\partial[M(y|X)]}{\partial X_k}\right]_{X=\bar{X}} \cdot \frac{\bar{X}_k}{M(y|\bar{X})} \quad , \tag{35}$$

where $M(y|X)$ is the conditional median of y. And indeed Goldberger (1968) makes the point that, if the regression equation is log-linear, (as with ECT = 0 in (12-A)) and u is assumed to be normally distributed, the exponential

---

[13]. Note, however, that *estimated* elasticities are evidently stochastic measures.

of the systematic part of the R.H.S. is the conditional median function of y : it follows that in this case of constant elasticity, the intuitive measure defined above coincides with (35).

More generally, it is not so easy to take into account the non-symmetric feature of the distribution function of y in nonlinear models. More importantly, we are typically interested in describing mean, not median, behavior : it is therefore useful to focus on another measure of central tendency of the conditional distribution such as the expected value. In this case (neglecting the observation subscripts used in previous sections), one should write :

$$\eta_k(E) = \left[\frac{\partial[E(y|X)]}{\partial X_k}\right]_{X=\bar{X}} \cdot \frac{\bar{X}_k}{E(y|\bar{X})} \quad . \tag{36}$$

It is well known that, if (33) is linear (with a constant term) or log-linear, the intuitive "sample" measure (34) coincides with the strict estimated "expected value" measure (36); in the log-linear case, the three measures defined above coincide. Both strict measures (35) and (36) have the further advantage of escaping the objection raised against the *stochastic* intuitive definition (34) : they are *fixed* because they are based on parameters of the distribution of y. The expected value elasticity has been promoted in a general way by various authors, such as Kmenta (1971); more specifically, it has also been advocated in the context in which BCT are used [Huang and Grawe (1980)]. Overall, any growth in its use has been even slower than that of the practice of reporting elasticities instead of regression coefficients. The difficulty of calculating expected values partly explains this state of affairs.

5.2. Elasticity with Respect to Variables which Contain Observed Zeros

All formulas just stated use $\bar{X}_k$, the sample mean, as the base level of interest for X_k and, for the dependent variables, various base levels also obtained from the whole sample. They implicitly assume that one is interested in the relative impact on y in general of X_k in general. We now wish to shift the focus to cases in which one is interested in the impact of X_k *when it happens* on y *in general*. In our applications, for instance, we might ask about the impact on transit demand of transit strike days or summer fairs when they happen. The question applies to explanatory variables which contains zeros because the activity which they describe does not occur all the time. In our applications, such variables are not transformed by a BCT[14]. To simplify the exposition, we will continue to assume that we are not interested in a sub-sample of y [15].

5.2.1. Quasi-dummy variables

Let us call variables such as these, which contain zero observations but which are not Boolean (0, or 1), *quasi-dummy* variables. What is the proper way to obtain elasticities with respect to quasi-dummy variables if one is not interested in the effect of the event in general but only in its effect when it occurs ? Consider first the case of intuitive elasticities. If we

---

[14]. Note that our L-1.1 algorithm does not provide for the possibility that quasi-dummy variables be affected by a BCT (with $\lambda \neq 0$). This is not because such a transformation would not be invariant to units of measurement —one could compensate for the effect on zero observations with a binary variable.

[15]. Clearly, base levels for the dependent variable and explanatory variables are arbitrary and could all be different.

replace, in (34) the mean over all observations, $\bar{X}_k$, by $\bar{X}_k^+$, the mean computed only from strictly positive observations, we obtain the desired formula :

$$\eta_k(S)^+ = \left[\frac{\partial y}{\partial X_k}\right]_{X=\bar{\bar{X}}} \cdot \frac{\bar{X}_k^+}{\bar{y}} \quad , \tag{37}$$

where $\bar{\bar{X}} = (\bar{X}_1, \ldots, \bar{X}_k^+, \ldots, \bar{X}_K)$ [16].

This correction of elasticities carries over[17] to the expected value case :

$$\eta_k(E)^+ = \left[\frac{\partial[E(y|X)]}{\partial X_k}\right]_{X=\bar{\bar{X}}} \cdot \frac{\bar{X}_k^+}{E(y|\bar{X})} \tag{38}$$

In (38), one might clearly prefer to evaluate the expected value at $\bar{\bar{X}}$ rather than at $\bar{X}$, if one felt that this would be more representative and would not care to have a common reference point for the expected value of y.

---

16. Note that the vector $\bar{\bar{X}}$ might contain other quasi-dummy variables besides X_k. When interest centers exclusively on the effect of X_k when the corresponding event occurs, the other dummy variables may be ascribed their usual $\bar{X}$ values.

17. Formally, for the model defined in (12), the expected value elasticity, written for any observation t, is

$$\eta_{kt}(E) = \frac{1}{E(y_t)} \int_{R(w)} [y_t(w)]^{1-\lambda_y} [X_{kt}^{\lambda_{xk}} \beta_k + H_{kt}(w)] \, N(w) \, dw$$

where $N(w) = f(w;\sigma^2)$ in (18), $E(y_t)$ and $y_t(w)$ are defined in (18), R(w) is the domain of integration for w (see Liem et al. (1983) for details) and $H_{kt}(w)$ equals 0 if X_k is not a Z_m variable and otherwise equals

$$\frac{1}{2} Z_{mt}^{\lambda_{zm}} \delta_m \sqrt{f(Z_t)} \left[\sum_\ell \rho_\ell \left(y^*_{t-\ell} - \sum_k X^*_{k,t-\ell} \beta_k\right) + v\right] , \text{ where } X^*_{kt} = \frac{X_{kt}^{(\lambda_{xk})}}{\sqrt{f(Z_t)}} .$$

5.2.2. Dummy variables

Dummy variables pose a special problem because the partial derivative $\partial(\cdot)/\partial X_k$ used in both the intuitive "sample" elasticity (34) and the strict "expected value" elasticity (36) does not exist for binary variables. In this case, one has to go back to the discrete (arc) measure of sensitivity and find the percentage variation of the dependent variable (or of its expected value) due to the dummy variable.

To simplify the discussion, let us start again with the intuitive (sample) notion, assuming that the elasticity will be evaluated at the means.

Define $\bar{\hat{y}}_1$ as the mean value of y obtained from a model evaluated with D, the dummy variable, set at its non-zero value (usually one) for all observations and $\bar{\hat{y}}_o$ as the mean value of y obtained from the same model evaluated with the dummy variable set at zero for all observations. Denote by D_1 the numerical value of the dummy variable when the event which it represents occurs (usually one), and by D_o its numerical value otherwise (zero). Then four cases can be distinguished, according to whether one is interested in :

- removing the effect of the dummy and evaluating the % change of y

$$\text{at } y_o : \quad \left[\frac{\bar{\hat{y}}_1 - \bar{\hat{y}}_o}{\bar{\hat{y}}_o}\right] \div (D_1 - D_o) \quad , \tag{39-A}$$

$$\text{or at } y_1 : \quad \left[\frac{\bar{\hat{y}}_1 - \bar{\hat{y}}_o}{\bar{\hat{y}}_1}\right] \div (D_1 - D_o) \quad ; \tag{39-B}$$

- or in adding the effect of the dummy and evaluating the % change of y

$$\text{at } y_o : \quad \left[\frac{\bar{\hat{y}}_o - \bar{\hat{y}}_1}{\bar{\hat{y}}_o}\right] \div (D_o - D_1) \quad , \tag{39-C}$$

$$\text{or at } y_1 : \quad \left[\frac{\bar{\hat{y}}_o - \bar{\hat{y}}_1}{\bar{\hat{y}}_1}\right] \div (D_o - D_1) \quad . \tag{39-D}$$

The last two cases obviously reduce to the first two and can be neglected.

Moreover, evaluation at y_o would overstate the impact of a dummy with a positive coefficient on the dependent variable in general and evaluation at y_1 understate it (and conversely for a dummy with a negative coefficient). Evaluation at the midpoint would yield the same result in both directions :

$$\left[\frac{(\bar{\hat{y}}_1 - \bar{\hat{y}}_o)\ 2}{(\bar{\hat{y}}_1 + \bar{\hat{y}}_o)}\right] \div (D_1 - D_o) \quad . \tag{40}$$

One could also envisage evaluating the sensitivity at the sample mean $\bar{y}$ to remain as close as possible to the continuous case :

$$\left[\frac{\bar{\hat{y}}_1 - \bar{\hat{y}}_o}{\bar{y}}\right] \div (D_1 - D_o) \tag{41}$$

Blaylock and Smallwood (1983) have developed the special case (39-A) with a BCT on y and found :

$$\left[\frac{\hat{y}_1 - \hat{y}_o}{\hat{y}_o(D_1 - D_o)}\right]_{\bar{\hat{y}}_1, \bar{\hat{y}}_o} = \begin{cases} \left\{\left\{\left[1 + \frac{\lambda_y \beta_d}{\hat{y}_o^{\lambda_y}}\right]^{1/\lambda_y} - 1\right\}\right\}_{\bar{\hat{y}}_o, \hat{\lambda}_y, \hat{\beta}_d} & , \lambda_y \neq 0, \quad (39\text{-A1}) \\ \left\{\left[\exp(\beta_d)\right] - 1\right\}_{\hat{\beta}_d} & , \lambda_y = 0 \quad (39\text{-A2}) \end{cases}$$

where β_d is the regression coefficient associated with the dummy variable, the contents of the square parentheses in (39-A1) must be positive to make the computation of the inverse power $1/\lambda_y$ feasible, and $\hat{y}_o$ is obtained by "unrolling" the BCT on y. We have pointed out in Section 3.4 that such an operation is not trivial and may be impossible, especially when $\lambda_y < 0$. In addition, this formulation, like all variants of (39) and (40) - (41), is not stated in terms of expected values.

Proper evaluation of the expected value formulation of any of the 4 distinct measures of sensitivity defined above requires, for the numerator, calculating the difference between two expected values[18] and, for the denominator, obtaining corresponding estimates for the reference level of the dependent variable :

$$\eta_k(E)^d = \frac{E(y|X)_{X=\tilde{X}_1} - E(y|X)_{X=\tilde{X}_o}}{\text{reference level}} \quad , \text{ where k refers to the binary variable,} \quad (42)$$

---

[18]. In effect, the numerator is equal to

$$\int_{R(w)} y(w_t)_{X=\tilde{X}_1} f(w;\sigma^2)\, dw - \int_{R(w)} y(w_t)_{X=\tilde{X}_o} f(w;\sigma^2)\, dw$$

where the domain of integration is $[w_t^{**}, \infty]$ for $\lambda_y > 0$, $[-\infty, +\infty]$ for $\lambda_y = 0$ and $[-\infty, w_t^{*}]$ for $\lambda_y < 0$.

and where

$$\tilde{X}_1 = (\bar{X}_1,\ldots,\bar{X}_{k-1},\ 1,\ \bar{X}_{k+1},\ldots,\bar{X}_k)\ ,$$

$$\tilde{X}_o = (\bar{X}_1,\ldots,\bar{X}_{k-1},\ 0,\ \bar{X}_{k+1},\ldots,\bar{X}_k)\ ,$$

and the reference levels are :

1. $E(y|X)_{X=\bar{X}}$ for (41) [19], (42-1)

2. $\frac{1}{2}\left[E(y|X)_{X=\tilde{X}_1} + E(y|X)_{X=\tilde{X}_o}\right]$ for (40), (42-2)

3. $E(y|X)_{X=\tilde{X}_1}$ for (39-B), (42-3)

4. $E(y|X)_{X=\tilde{X}_o}$ for (39-A). (42-4)

Depending on the value of λ_y, expected values can be calculated from (20), (21) or (22) after redefining $y(w_t)$ as either $y(w_t)_1$ or $y(w_t)_o$ to reflect the fact that, in the former case, $y(w_t)$ is computed with D = 1 and, in the latter, with D = 0. Assuming that the probability of being at the limit is negligible[20], we obtain for (42), in the special linear ($\lambda_y = 1$) and logarithmic cases ($\lambda_y = 0$), the values shown in Table 2 where n_o is the number of observations over which the dummy variable is equal to zero, n_1 is the number of observations over which it is equal to 1 and N is the total number of observations considered.

---

19. Clearly, this case is the analog of (38). If we had used $E(y|\bar{\bar{X}})$ in (38), then (42-3) would be the corresponding expression.

20. If that were not true, the expressions for the linear case would be more complicated. For instance, case 1.A in Table 2 would be :

$$\frac{\beta_d \int_{w_t^{**}}^{\infty} f(w;\sigma_w^2)\ dw}{\int_{w_t^{**}}^{\infty} y(w_t)\ f(w;\sigma_w^2)\ dw}\ .$$

TABLE 2

SPECIAL CASES OF THE EXPECTED VALUE ELASTICITY FOR BINARY VARIABLES

Reference level \ Case	*Linear ($\lambda_y=1$) A	Logarithmic ($\lambda_y=0$) B
1. $E(y\|X)_{X=\bar{X}}$	$\frac{\beta_d}{\bar{y}}$	$\frac{e^{\beta_d}-1}{\beta_d(n_o/N)}$
2. $\frac{E(y\|X)_{X=\tilde{X}_1} + E(y\|X)_{X=\tilde{X}_o}}{2}$	$\frac{2\beta_d}{2\bar{y} + \beta_d(\frac{n_o - n_1}{N})}$	$\frac{2(e^{\beta_d}-1)}{e^{\beta}+1}$
3. $E(y\|X)_{X=\tilde{X}_1}$	$\frac{\beta_d}{\bar{y} + \frac{n_o\beta_d}{N}}$	$\frac{2(e^{\beta_d}-1)}{e^{\beta}}$
4. $E(y\|X)_{X=\tilde{X}_o}$	$\frac{\beta_d}{\bar{y} - \frac{n_1\beta_d}{N}}$	$e^{\beta_d}-1$

* It is assumed that the probability of being at the lower limit is negligible

For the special cases shown in Table 2, the intuitive and expected value elasticities will coincide. For instance the Blaylock and Smallwood intuitive form (39-A1) yields $\beta_d/\bar{y}_o$, which is identical to case 4.A; case 4.B is identical to (39-A2) which also coincides with the special case formulated by Halvorsen and Palmquist (1980). Generally speaking, however, there is no way to avoid integrating as in footnote 18 if one wants to obtain one of the four variants of (42).

5.2.3. Approximation used and reporting format

To avoid at this time the additional computational burden of using (38) for quasi-dummy variables and (42-1) for dummy variables, we used instead, for both,

$$\eta_k(E) \cdot \frac{\bar{X}_k^+}{\bar{X}_k} \quad , \tag{43}$$

an approximation obtained simply by multiplying the strict expected value elasticity obtained for the whole sample by the ratio of the mean value of X_k computed over positive observations to the mean computed over all observations. For quasi-dummy variables, the value of this approximation depends on the appropriateness of evaluating the derivative in (38) at $\bar{X}$ instead of $\bar{\bar{X}}$. In our case, given that our algorithm scales variables to the order of magnitude of 1, it can be shown that the approximation is reasonable, especially if the regression coefficient obtained is small[21].

---

[21]. For example, in the simpler intuitive case, the partial derivative $(\partial y/\partial X_k) = (\beta_k/y^{\lambda_y - 1})$ can itself be derived with respect to X_k to verify this point. The result, $\partial(\partial y/\partial X_k)/\partial X_k = -\beta_k^2 (\lambda_y - 1)/y^{(2\lambda_y - 1)}$, has a denominator of the order of 1 because of our scaling of y described in the Appendix and a numerator which is small, for small values of $|\lambda_y - 1|$, if $|\beta_k| < 1$. In Table 3, $|\lambda_y - 1|$ as well as the coefficients of variables STR, FIR, EX7, EX8, EX9, EX0, EX1, ZA and ZSC are all smaller than 1 in absolute value except that of ZSC which is of the order of 1.5.

For dummy variables, using (43) instead of (42-1) involves a more complicated error[22]. In the linear case, (43) yields an *exact* result, namely case 1.A of Table 2, if the probability of being at the limit is negligible.

In the tables of results to be presented shortly, we will use the following convention to distinguish among the ways of computing expected value elasticities :

- a code name which is not underlined means that we are applying (36) and using all available observations on X_k to compute the elasticity for that variable;
- a code name underlined once means that we are applying (43) and using only the positive values of X_k to compute the elasticity for that variable;
- a code name underlined twice means that we are applying (43) to a dummy variable and using only the positive values of that variable to compute the elasticity.

(Code 3)

---

[22]. The numerators of the two expressions differ. In (43), we have $\beta_k \int_{R(w)} [y_t(w)]^{1-\lambda_y} N(w)\, dw$; in (42-1), the numerator is $\int_{R(w)} [y_t(w)_1] N(w)\, dw - \int_{R(w)} [y_t(w)_o] N(w)\, dw$. In the intuitive case, the values for $\lambda_y = 0$ are respectively β_o and $e^{\beta_d} - 1$. Considering that the first two terms of the Taylor expansion of e^{β_d} are $1 + \beta_d$, the approximation is acceptable if $|\beta_d| < 1$. In Table 3, the coefficients of the dummy variables STRUCT1, CHEAT2 and CARDEX, which have a positive value for several observations (ANT and ANTL1 are defined only for one observation each), are between 0.02 and 0.18.

5.3. Elasticities

We will now examine complete regression results for some representative experiments taken from Table 1. The first set, shown in Table 3, shows the impact of adding AU, and then HG_1, to a model which has the same BCT on both the dependent and the (strictly positive) independent variables : experiments 2.0, 2.1 and 2.3 from series A-1, S-1 and G-1 are used, in part to maintain comparability with previously published results. The second set, shown in Table 4, is chosen to illustrate the impact of pushing the specification (12) to the point of diminishing returns (on the log-likelihood) : nine cases from series S-4 of Table 1 are shown and the most general one has 3 BCT, 12 autoregressive parameters and 2 heteroscedasticity variables considered simultaneously with the other parameters of the model.

Table 3 and Table 4 contain three sections. The first section shows, for a given experiment (column), the elasticity of the dependent variable with respect to each X_k (whose code name is presented according to (Code 3)). Under each elasticity value, the asymptotic t-statistic of the β_k used to calculate that elasticity is found in parentheses[23]. Under the t-statistics,

---

[23]. We have pointed out in Section 3.3 that the t-values associated with the coefficients of X_k and Z_m were not invariant to units of measurement when a BCT was estimated. See also the Appendix on this point.

i) the absence of indication means that the variable X_k was not transformed by a BCT because it was not strictly positive;

ii) the indication FL1 means that the λ of the BCT associated with this variable is fixed to a particular value common to all X_k which belong to group 1; the fixed value for each group is found in the second section of the table;

iii) the indication EL1 means that the λ of the BCT associated with this variable is estimated and is common to all X_k which belong to group 1; the estimated value for each group is found in the second section of the table.

The second section of each table contains

a) estimates of the ρ_ℓ autocorrelation parameters and their asymptotic t-tests;

b) fixed or estimated values of the BCT on the X_k and on the dependent variable y; under the estimated values, the first parenthesis presents the asymptotic t-tests under the hypothesis that the parameter is equal to 0 (logarithmic case); in the second parenthesis, the t-test is made under the assumption that the parameter is equal to 1 (linear case);

c) estimated values for the δ_m and the BCT on the Z_m heteroscedasticity variables; for the BCT, the two t-statistics in parentheses have the same meaning as in (b).

The third section contains log-likelihood values, the pseudo-(L)-R^2 described by (32) and other information on the experiment in question.

5.3.1. The three demand equations

In Table 3, one can see that the stochastic specification has a considerable impact on the sign and size of many of the elasticities, even in some cases where the t-statistics are rather large.

Consider first the variables whose anticipated sign is uncertain. For instance, other goods are a complement to urban travel (with positive cross-elasticity of price (P)) in some cases but a substitute (with negative price cross-elasticity) in other cases of the S and G equations; similarly, the influence of cold and warm temperature TLK depends on the stochastic specification of the A and G equations. The influence of snowfall SFK and college and university activity RCUTP[24] varies in A. For RCUTP, the t-statistics are superior to 1.5 in absolute value. Moreover, the influence of congestion (T2T) on the demand for gasoline, which would be positive if increased rates of consumption per unit of distance are not compensated by reduced total kilometers of travel and negative if they are, varies in size and sign among the three cases of equation G. The same is true for the aggregation variable SHD and the summer fair of 1971, EX1.

Now examine the variables whose sign is anticipated to be positive or negative. The following generalizations can be made with respect to changes in signs as one successively adds autocorrelation (column 2.1) and then heteroscedasticity of a general nature (column 2.2) to the basic specification (column (2.0) :

---

[24]. We do not know whether university attendance reduces other social activities of students sufficiently to imply a net gain or loss.

TABLE 3

ELASTICITIES AND OTHER STATISTICS : SELECTED CASES FROM SERIES A-1, S-1 AND G-1 OF TABLE 1

		Adults			Schoolchildren			Gasoline		
	CODE NO. *	2.0	2.1	2.3	2.0	2.1	2.3	2.0	2.1	2.3
	DEP.VAR. *	XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
P * PRICES		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
REAL MUCTC STUDENT FARE	FSCD				-.292 (-1.04) EL1	-.419 (-1.75) EL1	-.410 (-1.88) EL1			
REAL MUCTC ADULT TRANSIT FARE	FAD	-.121 (-1.08) EL1	-.188 (-3.07) EL1	-.193 (-2.91) EL1				.004 (.03) EL1	.081 (.85) EL1	.097 (.71) EL1
REAL PRICE OF REGULAR GASOLINE	AUMID							-.158 (-.89) EL1	-.008 (-.04) EL1	-.055 (-.27) EL1
CONSUMER PRICE INDEX, MONTREAL	P	.189 (.43) EL1	.101 (.43) EL1	.111 (.48) EL1	-1.800 (-.74) EL1	.578 (.30) EL1	-.186 (-.11) EL1	-1.540 (-1.27) EL1	.782 (.85) EL1	.672 (.76) EL1
M-Q * MOTOR VEHICLE - QUANTITY		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
CARS IN MUCTC AREA	NC	-.318 (-2.25) EL1	-.388 (-3.28) EL1	-.393 (-3.38) EL1	-.710 (-.78) EL1	-1.140 (-1.68) EL1	-1.040 (-1.64) EL1	.421 (1.42) EL1	.117 (.30) EL1	-.055 (-.15) EL1
N-T * NETWORK - TIME,SERVICE LEVELS		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
WAIT TIME FOR MUCTC VEHICLES	TW	-.307 (-3.74) EL1	-.044 (-.86) EL1	-.013 (-.20) EL1	-.392 (-.85) EL1	-.277 (-.74) EL1	-.331 (-.74) EL1	.557 (2.28) EL1	-.036 (-.15) EL1	-.098 (-.39) EL1
MUCTC IN-VEHICLE TRAVEL TIME	TT	-.496 (-8.13) EL1	-.490 (-15.08) EL1	-.486 (-14.60) EL1	.513 (.69) EL1	-.102 (-.25) EL1	-.057 (-.09) EL1	.082 (.78) EL1	.202 (1.04) EL1	.187 (1.04) EL1
OFF-PEAK TRAVEL TIME BY CAR, MTL	T2T	-.323 (-1.19) EL1	.082 (.59) EL1	.127 (1.06) EL1	1.940 (1.24) EL1	.710 (.58) EL1	.818 (.54) EL1	.743 (1.65) EL1	-.082 (-.14) EL1	.030 (.08) EL1
COMPLETE STRIKES BY MUCTC	STR ---	-.555 (-5.42)	-.538 (-15.74)	-.539 (-14.80)	-.682 (-3.49)	-.587 (-4.55)	-.550 (-4.71)	.107 (1.21)	.054 (1.81)	.059 (1.68)
RESTRUCTURATION OF LINES FOR METRO	STRUCT1 *******				.072 (.44)	-.068 (-.68)	-.058 (-.45)			
N-I * NETWORK - INFRASTRUCTURE,WEATHER		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
TOO HOT OR COLD RELATIVE TO 58 F.	TLK	-.001 (-.17)	.001 (.07)	-.002 (-.23)	.057 (1.82)	.068 (1.86)	.020 (.65)	-.021 (-1.12)	.014 (.83)	.001 (.06)
RAINFALL PER DAY, DORVAL AIRPORT	RFK	-.011 (-1.95)	-.008 (-2.84)	-.008 (-2.77)	-.014 (-.80)	-.013 (-1.02)	-.011 (-1.14)	-.007 (-.67)	-.010 (-1.07)	-.010 (-1.17)
SNOWFALL PER DAY, DORVAL AIRPORT	SFK	-.004 (-1.43)	.000 (.22)	-.000 (-.03)	-.028 (-1.96)	-.013 (-1.40)	-.009 (-1.08)	.013 (1.48)	.007 (1.12)	.007 (1.17)
ACCUMULATED SNOWFALL,DORVAL AIRPORT	CSFHR	.019 (3.55)	.002 (.89)	.003 (.77)	.018 (1.13)	.008 (.58)	.013 (.91)	-.036 (-3.47)	-.010 (-.97)	-.007 (-.61)
Y-G * CONSUMERS - GENERAL CHARACTERISTICS		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
REAL AVERAGE WEEKLY EARNINGS, MTL	RAWEM	.601 (1.68) EL1	.588 (3.10) EL1	.578 (3.10) EL1	.214 (.27) EL1	.270 (.43) EL1	.823 (.97) EL1	.012 (.03) EL1	-.108 (-.18) EL1	-.256 (-.42) EL1

. . . /

TABLE 3 (continued)

Y-A • CONSUMERS - AGE		XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
COMPUTER CONTROL REDUCED FARE CARDS	CHEAT2 ******				-.083 (-1.29)	-.019 (-.28)	-.057 (-.55)			
SUMMER EXPIRAT.REDUCED FARE CARDS	CARDEX ******				.014 (.19)	-.035 (-.52)	-.029 (-.42)			
A • ACTIVITY FINAL AND INTERMEDIATE		**XCANFA**	**XCANFA**	**XCANFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**VNEM**	**VNEM**	**VNEM**
INDEX SCHOOL,JR.COLL.ACTIVITY,MTL	RSCMP				.474 (4.42)	.370 (3.99)	.341 (4.05)			
REAL RETAIL SALES INDEX	SHONF	.083 (5.35) EL1	.050 (1.82) EL1	.044 (1.34) EL1	.125 (2.80) EL1	.112 (2.34) EL1	.091 (1.57) EL1	.214 (8.01) EL1	.178 (2.39) EL1	.139 (1.88) EL1
TREND SECOND.SCHOOL REGISTRATION	TEND				.805 (1.73) EL1	.803 (1.88) EL1	.588 (1.38) EL1			
EMPLOYMENT ACTIVITY INDEX,MTL	EANFP	.134 (2.83) EL1	.362 (4.83) EL1	.388 (5.24) EL1						
EMPLOYMENT INDEX,MTL	EI							1.480 (3.30) EL1	.218 (.63) EL1	.233 (.65) EL1
INDEX COLL.,UNIV.ACTIVITY,MTL	RCUTP	.017 (2.14)	-.020 (-1.52)	-.022 (-1.68)						
INDEX SCHOOL,COLL.,UNIV.ACTIV.,MTL	TRCUSP							-.071 (-2.34) EL1	-.022 (-1.04) EL1	-.159 (-.29) EL1
VACATION INDEX	VWN							.010 (1.58)	.005 (.53)	.008 (.80)
EXPO 87	EX7 ---	.318 (3.13)	.348 (9.21)	.350 (8.68)	.232 (3.45)	.188 (3.45)	.189 (3.04)	.119 (1.45)	.059 (.89)	.044 (.77)
MAN AND HIS WORLD 1988	EX8 ---	.089 (2.85)	.086 (4.63)	.087 (4.52)	.068 (1.32)	.058 (1.31)	.048 (.83)	.091 (1.23)	.057 (.90)	.047 (.73)
MAN AND HIS WORLD 1989	EX9 ---	.080 (2.09)	.050 (2.17)	.049 (2.11)				.072 (1.09)	.041 (.50)	.031 (.43)
MAN AND HIS WORLD 1970	EX0 ---	.038 (1.20)	.009 (.32)	.007 (.25)				-.042 (-.55)	-.025 (-.25)	-.033 (-.36)
MAN AND HIS WORLD 1971	EX1 ---	.037 (.90)	.003 (.08)	.000 (.01)				-.011 (-.10)	.005 (.05)	-.010 (-.09)
ET-AD • ET CETERA - ADMINISTRATIVE		**XCANFA**	**XCANFA**	**XCANFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**VNEM**	**VNEM**	**VNEM**
MULT.FARE ZONES SCH.CHILD.DBL.COUNT	ZSC ---				.107 (1.43)	.090 (2.23)	.095 (2.32)			
MULT.FARE ZONES ADULT DBL.COUNTING	ZA --	.018 (1.34)	.036 (4.11)	.037 (4.08)						
SPECULATIVE TICKET PURCHASE	ANT ***	-.048 (0.00)	-.042 (-2.48)	-.041 (-2.73)	.289 (0.00)	.343 (1.03)	.419 (.98)			
SPECULATIVE TICKET PURCHASE LAG.1M.	ANTL1 *****	.035 (0.00)	.009 (.94)	.013 (1.30)	-.128 (0.00)	-.137 (-.67)	-.137 (-.38)			
ET-AG • ET CETERA - AGGREGATION		**XCANFA**	**XCANFA**	**XCANFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**VNEM**	**VNEM**	**VNEM**
WORKDAYS PER MONTH	WD	.531 (8.47) EL1	.760 (4.55) EL1	.805 (5.11) EL1	.814 (2.42) EL1	.083 (2.00) EL1	.715 (2.19) EL1	.353 (1.98) EL1	.243 (.79) EL1	.151 (.49) EL1
SATURDAYS PER MONTH	SAT	.073 (2.61) EL1	.114 (2.81) EL1	.123 (3.15) EL1	.125 (1.44) EL1	.055 (.84) EL1	.071 (.82) EL1	-.022 (-.33) EL1	-.001 (-.01) EL1	.000 (.01) EL1
SUNDAYS AND HOLIDAYS PER MONTH	SHD	.088 (2.52) EL1	.107 (2.32) EL1	.117 (2.84) EL1	.087 (1.02) EL1	.058 (.77) EL1	.088 (1.29) EL1	-.034 (-.88) EL1	.042 (.51) EL1	-.013 (-.18) EL1
ET-SC • ET CETERA - SEASONALITY,CONSTANT		**XCANFA**	**XCANFA**	**XCANFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**VNEM**	**VNEM**	**VNEM**
REGRESSION CONSTANT	CONSTANT	-.013 (-.03)	-1.070 (-3.15)	-1.170 (-3.23)	-3.190 (-1.89)	-1.210 (-1.20)	-1.450 (-1.41)	-.777 (-1.08)	1.180 (.82)	2.130 (1.38)

. . . /

TABLE 3 (end)

	CODE NO. *	2.0	2.1	2.3	2.0	2.1	2.3	2.0	2.1	2.3
	DEP.VAR. *	XCANFA	XCANFA	XCANFA	XCSCNFA	XCSCNFA	XCSCNFA	VNEM	VNEM	VNEM
AUTOCORRELATION										
	RHO 1		.014	.009		.138	.255		.193	.196
			(.29)	(.21)		(1.78)	(2.89)		(3.72)	(4.12)
	RHO 3					-.344	-.186			
						(-3.80)	(-1.74)			
	RHO 4					-.086	-.055			
						(-1.08)	(-.57)			
	RHO 12		.910	.922		.382	.471		.731	.735
			(19.01)	(21.64)		(4.52)	(6.38)		(14.62)	(18.19)
BOX-COX TRANSFORMATIONS										
ESTIMATED LAMBDA(X) - GROUP 1	EL1	.751	.921	.917	.511	.723	1.005	.855	.017	.048
		(2.17)	(7.85)	(7.27)	(2.67)	(3.27)	(4.80)	(4.49)	(.07)	(.16)
		(-.72)	(-.88)	(-.88)	(-2.58)	(-1.25)	(.02)	(-.78)	(-3.92)	(-3.42)
ESTIMATED LAMBDA(Y)	ELY	.751	.921	.917	.511	.723	1.005	.855	.017	.048
		(2.17)	(7.85)	(7.27)	(2.67)	(3.27)	(4.80)	(4.49)	(.07)	(.16)
		(-.72)	(-.88)	(-.88)	(-2.58)	(-1.25)	(.02)	(-.78)	(-3.92)	(-3.42)
HETEROSKEDASTICITY										
ASSOCIATED VARIABLE = SHONF	DELTA 1			.510						
				(.37)						
ASSOCIATED VARIABLE = TEND	DELTA 2						.119			
							(.48)			
ASSOCIATED VARIABLE = TRCUSP	DELTA 3									.037
										(.41)
ASSOCIATED VARIABLE = SHONF	LAM(E) 1			.191						
				(.05)						
				(-.23)						
ASSOCIATED VARIABLE = TEND	LAM(E) 2						1.074			
							(1.04)			
							(.07)			
ASSOCIATED VARIABLE = TRCUSP	LAM(E) 3									2.113
										(1.05)
										(.58)
LOG-LIKELIHOOD		272.589	308.739	309.758	1.399	28.203	37.480	28.563	48.980	53.010
PSEUDO-(L)-R2		.955	.970	.971	.911	.935	.942	.981	.989	.991
NUMBER OF OBSERVATIONS		188	188	188	188	188	188	188	188	188
ESTIMATION PERIOD		13-181	13-181	13-181	13-181	13-181	13-181	13-181	13-181	13-181
NUMBER OF INDEPENDENT VARIABLES		28	28	28	27	27	27	26	26	26

- for the two DEMTEC model equations A and S, adding AU *improves* the results as can be readily verified by comparing columns 2.0 and 2.1 for variables TT in S, T2T in A and STRUCT1 [25] and CARDEX in S. Comparing columns 2.1 and 2.2 shows that the further addition of HG_1 did not have a significant impact, even for the S equation where such an addition considerably increased the log-likelihood.
- for the demand for gasoline equation G, adding AU yields *less convincing* results, as a comparison of columns 2.0 and 2.1 shows for variables TW and RAWEM. Further addition of HG_1 raises the log-likelihood value by a considerable amount, as a comparison of colum 2.1 and column 2.3 results shows, but the unreasonable cross-elasticities for TW and RAWEM become even more unreasonable.

5.3.2. Pushing the model to its limits

The results of the second section of Table 4 have already been used in Sections 4.4.2 and 4.4.3 to point out the influence of heteroscedasticity on the structure of autocorrelation (columns 2.3.2 and 2.4) and of increasing the number of BCT on the structure of autocorrelation (columns 2.4 and 3.4). The purpose of showing the first section is the analysis of elasticities. The following observations should be made after rejecting the logarithmic models (columns 0.1 and 0.3) in favor of the remaining linear or BCT models (which do not generally depart significantly from the linear form) :

---

[25]. It is known that the integration of bus and subway lines was made to the benefit of adults; schoolchildren cannot have gained by the increased access times implied by the integrated network.

- the signs are remarkably stable;
- with linear models and models with one or more BCT, the presence of the first variable to explain heteroscedasticity (TEND) changes the sign of at least one of the service level variables TW, TT or T2T in an unanticipated way and the sign of the variable CHEAT2, which describes the better control of the distribution of reduced fare cards, in an anticipated way;
- taking heteroscedasticity into account has a relatively large impact on the absolute value of the elasticities of many of the variables : P, NC, RAWEM : this possibly illustrates the point made in Section 4.3.2 about the sensitivity of correlations among regressors to their transformation by the powerful general form for heteroscedasticity. The elasticities for variables which have highly significant t-statistics in the linear form seem to be proportionately less affected by the addition of HG_1 than the others.

5.4. Values of Time

The sensitivity of the sign and size of the elasticities of transit demand exhibited in Table 3 and Table 4 should be expected to carry over to the implicit prices of time, calculated in the usual way by asking what variation in service level will produce the same effect on ridership as a unit change in the fare or :

$$\text{price of time} = \frac{\partial y / \partial \text{time}}{\partial y / \partial \text{price}} \quad . \tag{44}$$

TABLE 4

ELASTICITIES AND OTHER STATISTICS : SELECTED CASES FROM SERIES S-4 OF TABLE 1

Schoolchildren

	CODE NO. =	0.1	0.3	1.1	1.3	2.1	2.3	2.3.2	2.4	3.4
	DEP.VAR. =	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
P = PRICES		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
REAL MUCTC STUDENT FARE	FSCD	-.270	-.270	-.372	-.322	-.352	-.324	-.359	-.638	-.425
		(-1.51)	(-1.68)	(-2.98)	(-3.48)	(-1.88)	(-2.02)	(-1.88)	(-3.09)	(-1.71)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
CONSUMER PRICE INDEX, MONTREAL	P	-.859	-1.530	.722	.081	.514	.105	.550	1.440	4.270
		(-.49)	(-1.12)	(.58)	(.12)	(.40)	(.15)	(.42)	(.90)	(2.88)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
M-Q = MOTOR VEHICLE - QUANTITY		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
CARS IN MUCTC AREA	NC	.007	-.008	-1.100	-.798	-1.030	-.804	-1.070	-2.310	-1.700
		(.02)	(-.02)	(-2.15)	(-2.12)	(-2.11)	(-2.12)	(-2.05)	(-5.27)	(-1.08)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
N-T = NETWORK - TIME,SERVICE LEVELS		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
WAIT TIME FOR MUCTC VEHICLES	TW	-.712	-.820	-.034	.007	-.138	.038	-.149	-.547	-.080
		(-1.88)	(-1.80)	(-.27)	(.07)	(-.71)	(.21)	(-.71)	(-1.82)	(-.25)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
MUCTC IN-VEHICLE TRAVEL TIME	TT	.035	.192	-.550	-.522	-.477	-.546	-.485	.090	-.311
		(.08)	(.37)	(-1.92)	(-1.87)	(-1.57)	(-1.88)	(-1.52)	(.39)	(-1.00)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
OFF-PEAK TRAVEL TIME BY CAR, MTL	T2T	1.010	1.100	.028	-.083	.317	-.117	.310	.506	1.200
		(.77)	(1.02)	(.05)	(-.18)	(.43)	(-.30)	(.40)	(.55)	(.58)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1
COMPLETE STRIKES BY MUCTC	STR	-.811	-.788	-.603	-.597	-.631	-.592	-.631	-.575	-.592
	---	(-20.40)	(-19.05)	(-14.19)	(-11.63)	(-5.42)	(-4.50)	(-5.11)	(-8.81)	(-4.02)
RESTRUCTURATION OF LINES FOR METRO	STRUCT1	.016	.063	-.122	-.091	-.104	-.099	-.098	.082	-.098
	*******	(.17)	(.58)	(-1.87)	(-1.13)	(-1.28)	(-1.08)	(-1.20)	(.98)	(-1.00)
N-I = NETWORK - INFRASTRUCTURE,WEATHER		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
TOO HOT OR COLD RELATIVE TO 58 F.	TLK	-.073	-.079	-.027	-.052	-.034	-.050	-.033	.018	-.085
		(-1.50)	(-1.85)	(-.73)	(-1.59)	(-.88)	(-1.48)	(-.83)	(.70)	(-2.45)
RAINFALL PER DAY, DORVAL AIRPORT	RFK	-.021	-.021	-.019	-.017	-.019	-.017	-.019	-.018	-.018
		(-1.22)	(-1.49)	(-1.57)	(-1.94)	(-1.47)	(-1.85)	(-1.40)	(-1.94)	(-1.88)
SNOWFALL PER DAY, DORVAL AIRPORT	SFK	-.012	-.005	-.017	-.006	-.018	-.006	-.015	-.002	.000
		(-1.21)	(-.52)	(-2.41)	(-.82)	(-1.72)	(-.88)	(-1.59)	(-.34)	(.07)
ACCUMULATED SNOWFALL,DORVAL AIRPORT	CSFNR	.019	.015	.024	.022	.023	.022	.022	.004	.020
		(1.19)	(1.12)	(1.97)	(2.44)	(1.44)	(1.73)	(1.40)	(.42)	(1.84)
Y-G = CONSUMERS - GENERAL CHARACTERISTICS		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
REAL AVERAGE WEEKLY EARNINGS, MTL	RAWEM	-.259	.289	.394	.513	.186	.504	.215	1.410	.445
		(-.39)	(.41)	(.85)	(.98)	(.31)	(1.08)	(.39)	(3.61)	(.54)
		FL1	FL1	FL1	FL1	EL1	EL1	EL1	EL1	EL1

. . . /

TABLE 4 (continued)

Y-A = CONSUMERS - AGE		XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
COMPUTER CONTROL REDUCED FARE CARDS	CHEAT2 ******	-.113 (-2.77)	-.180 (-3.89)	-.000 (0.00)	-.042 (-.81)	-.003 (-.04)	-.042 (-.58)	-.004 (-.07)	-.083 (-1.16)	.031 (.38)
SUMMER EXPIRAT.REDUCED FARE CARDS	CARDEX ******	-.084 (-.85)	-.037 (-.62)	-.037 (-.75)	-.034 (-.81)	-.040 (-.77)	-.034 (-.78)	-.036 (-.69)	.032 (.83)	-.078 (-1.85)
A = ACTIVITY FINAL AND INTERMEDIATE		**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**
INDEX SCHOOL.JR.COLL.ACTIVITY.MTL	RSCMP	.388 (5.85)	.399 (7.33)	.285 (8.11)	.287 (8.80)	.281 (4.54)	.281 (4.62)	.283 (4.25)	.384 (7.18)	.289 (4.39)
REAL RETAIL SALES INDEX	SHONF	.059 (.99) FL1	.055 (.97) FL1	.094 (2.69) FL1	.090 (2.59) FL1	.089 (2.34) EL1	.091 (2.65) EL1	.095 (2.22) EL1	.121 (2.90) EL1	.184 (2.86) EL2
TREND SECOND.SCHOOL REGISTRATION	TEND	.311 (1.87) FL1	.281 (1.92) FL1	.557 (3.24) FL1	.444 (2.75) FL1	.585 (2.47) EL1	.440 (2.12) EL1	.588 (2.22) EL1	.809 (3.94) EL1	.171 (1.13) EL2
EXPO 67	EX7 ---	.180 (3.51)	.190 (3.24)	.153 (3.14)	.144 (2.92)	.153 (2.52)	.144 (2.41)	.149 (2.38)	.039 (.88)	.067 (1.85)
MAN AND HIS WORLD 1968	EX8 ---	.031 (.85)	.040 (.88)	.040 (.99)	.023 (.49)	.039 (.97)	.024 (.49)	.039 (.95)	.022 (.24)	.025 (.88)
ET-AD = ET CETERA - ADMINISTRATIVE		**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**
MULT.FARE ZONES SCH.CHILD.DBL.COUNT	ZSC ---	.057 (1.94)	.066 (2.47)	.087 (5.04)	.083 (4.78)	.082 (2.72)	.083 (2.78)	.083 (2.65)	.078 (2.50)	.104 (2.87)
SPECULATIVE TICKET PURCHASE	ANT ***	.286 (3.12)	.279 (2.69)	.404 (4.38)	.421 (4.21)	.372 (2.40)	.433 (2.11)	.388 (2.29)	.289 (2.55)	.436 (2.27)
SPECULATIVE TICKET PURCHASE LAG.1M.	ANTL1 *****	-.196 (-1.44)	-.186 (-.82)	-.248 (-3.73)	-.228 (-3.15)	-.243 (-2.93)	-.230 (-2.41)	-.243 (-2.82)	-.125 (-1.27)	-.130 (-1.08)
ET-AG = ET CETERA - AGGREGATION		**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**
WORKDAYS PER MONTH	WD	.338 (1.17) FL1	.451 (1.85) FL1	.331 (1.70) FL1	.373 (2.08) FL1	.334 (1.64) EL1	.367 (2.05) EL1	.338 (1.81) EL1	.791 (2.93) EL1	.334 (1.07) EL1
SATURDAYS PER MONTH	SAT	-.011 (-.15) FL1	.003 (.03) FL1	-.038 (-.71) FL1	-.014 (-.27) FL1	-.033 (-.59) EL1	-.015 (-.30) EL1	-.034 (-.80) EL1	.033 (.51) EL1	-.010 (-.19) EL1
SUNDAYS AND HOLIDAYS PER MONTH	SHD	.051 (.82) FL1	.089 (1.17) FL1	-.031 (-.54) FL1	-.007 (-.13) FL1	-.021 (-.38) EL1	-.011 (-.20) EL1	-.018 (-.30) EL1	.127 (1.68) EL1	-.036 (-.61) EL1
ET-SC = ET CETERA - SEASONALITY,CONSTANT		**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**	**XCSCNFA**
REGRESSION CONSTANT	CONSTANT	-.896 (-.40)	-2.350 (-1.25)	.032 (.07)	.007 (.02)	-.003 (0.00)	.018 (.04)	-.043 (-.07)	-2.290 (-2.75)	-1.140 (-2.17)

. . . /

TABLE 4 (end)

CODE NO. =	0.1	0.3	1.1	1.3	2.1	2.3	2.3.2	2.4	3.4
DEP.VAR. =	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA	XCSCNFA
AUTOCORRELATION									
RHO 1	.198	.106	.267	.150	.244	.155	.250	.371	-.022
	(1.83)	(.99)	(2.63)	(1.48)	(2.25)	(1.44)	(2.20)	(3.81)	(-.20)
RHO 2	-.198	-.122	-.134	-.082	-.158	-.078	-.153	.154	-.178
	(-1.78)	(-.99)	(-1.11)	(-.87)	(-1.32)	(-.64)	(-1.27)	(1.55)	(-1.45)
RHO 3	-.259	-.273	-.389	-.417	-.370	-.418	-.371	-.482	-.581
	(-2.24)	(-2.20)	(-3.26)	(-3.68)	(-3.28)	(-3.71)	(-3.25)	(-4.28)	(-5.23)
RHO 4	-.203	-.237	-.233	-.258	-.249	-.249	-.250	-.225	-.378
	(-1.85)	(-2.08)	(-2.33)	(-2.14)	(-2.39)	(-2.08)	(-2.28)	(-1.88)	(-3.15)
RHO 5	-.101	-.105	-.101	-.131	-.120	-.127	-.113	.217	-.308
	(-.93)	(-.87)	(-.85)	(-1.13)	(-1.01)	(-1.11)	(-.93)	(1.83)	(-2.55)
RHO 6	-.180	-.183	-.138	-.180	-.155	-.178	-.150	.038	-.405
	(-1.48)	(-1.65)	(-1.35)	(-1.87)	(-1.65)	(-1.82)	(-1.45)	(.28)	(-3.83)
RHO 7	-.389	-.422	-.317	-.372	-.338	-.384	-.340	-.415	-.477
	(-4.88)	(-4.45)	(-3.95)	(-4.18)	(-4.25)	(-4.07)	(-4.21)	(-3.57)	(-4.27)
RHO 8	-.068	-.145	-.062	-.093	-.070	-.089	-.089	-.024	-.219
	(-.83)	(-1.26)	(-.81)	(-.85)	(-.87)	(-.81)	(-.88)	(-.20)	(-1.98)
RHO 9	-.289	-.188	-.417	-.342	-.417	-.345	-.410	.184	-.454
	(-2.63)	(-1.53)	(-4.24)	(-3.83)	(-4.10)	(-3.87)	(-4.01)	(1.55)	(-4.58)
RHO 10	.057	-.054	.129	-.055	.099	-.050	.105	-.071	-.205
	(.88)	(-.53)	(1.36)	(-.59)	(1.03)	(-.53)	(1.04)	(-.52)	(-2.10)
RHO 11	-.131	-.072	-.124	-.048	-.140	-.040	-.145	-.179	-.223
	(-1.30)	(-.68)	(-1.25)	(-.48)	(-1.45)	(-.39)	(-1.48)	(-1.30)	(-1.89)
RHO 12	.284	.175	.191	.138	.190	.140	.193	.321	-.024
	(3.88)	(1.83)	(1.93)	(1.42)	(1.88)	(1.41)	(1.87)	(3.17)	(-.24)
BOX-COX TRANSFORMATIONS									
FIXED LAMBDA(X) - GROUP 1 FL1	0.000	0.000	1.000	1.000					
FIXED LAMBDA(Y) FLY	0.000	0.000	1.000	1.000					
ESTIMATED LAMBDA(X) - GROUP 1 EL1					.818	1.058	.804	.487	4.181
					(4.24)	(4.62)	(3.93)	(4.13)	(5.08)
					(-.95)	(.24)	(-.98)	(-4.71)	(3.85)
ESTIMATED LAMBDA(X) - GROUP 2 EL2									1.058
									(3.05)
									(.17)
ESTIMATED LAMBDA(Y) ELY					.818	1.058	.804	.487	.885
					(4.24)	(4.62)	(3.93)	(4.13)	(3.52)
					(-.95)	(.24)	(-.98)	(-4.71)	(- 55)
HETEROSKEDASTICITY									
ASSOCIATED VARIABLE = TEND DELTA 1		.013		.048		.049		.043	.060
		(.38)		(.58)		(.59)		(.43)	(.80)
ASSOCIATED VARIABLE = SHONF DELTA 2							-.071	-.055	-.034
							(-.08)	(-1.39)	(-.40)
ASSOCIATED VARIABLE = TEND LAM(E) 1		2.085		1.577		1.561		1.672	1.493
		(1.72)		(1.87)		(1.97)		(1.57)	(1.95)
		(.89)		(.72)		(.71)		(.83)	(.85)
ASSOCIATED VARIABLE = SHONF LAM(E) 2							.842	4.617	3.806
							(.04)	(7.43)	(1.80)
							(-.01)	(5.82)	(1.33)
LOG-LIKELIHOOD	30.881	38.705	45.703	58.982	46.145	58.993	46.178	64.808	68.079
PSEUDO-(L)-R2	.937	.943	.947	.954	.948	.954	.948	.958	.960
NUMBER OF OBSERVATIONS	189	189	189	189	189	189	189	189	189
ESTIMATION PERIOD	13-181	13-181	13-181	13-181	13-181	13-181	13-181	13-181	13-181
NUMBER OF INDEPENDENT VARIABLES	27	27	27	27	27	27	27	27	27

The result of this computation for adults and schoolchildren is found in Table 5 for the cases shown in Table 3.

TABLE 5

IMPLICIT VALUES OF TIME, ADULTS AND SCHOOLCHILDREN ($/hour)

	ADULTS			SCHOOLCHILDREN		
* Experiment #	2.0	2.1	2.3	2.0	2.1	2.3
Transit Wait-time TW	6.30	0.59	0.16	1.21	0.59	0.73
Transit In-vehicle time TT	1.90	1.23	1.17	-0.30	0.04	0.02

* All cases are from Table 3.

In interpreting these results, we have to remember that the equation for the adult market contains no variable to account for the impact of changed access times due to the addition of the subway to the bus network and the simultaneous restructuring of the lines to feed the subway. For that reason, the in-vehicle travel time variable reflects that structural change (including modified comfort) as well as the influence of modified speeds. Despite that, it is clear that the relative price of wait-time falls too much for the adult market when autocorrelation and heteroscedasticity are taken into account in columns 2.1 and 2.3, respectively. Results are more reasonable in the schoolchildren equation except for the negative sign in column 2.0. If we keep in mind that the average manufacturing wage during

our sample period was about 2 $ per hour in constant 1956 dollars, it is clear that one could not use these implicit prices without further work centered on an examination on the robustness of these results.

6. CONCLUSION

In this paper we have :

i) provided a rationale for a Box-Cox regression model with heteroscedasticity and multiple autocorrelation. Our approach consists of viewing the problem *a priori* as a two-limit Tobit applied to a sample containing no limit observations and of calculating *a posteriori* the appropriateness of that assumption;

ii) shown the importance of making a correction for heteroscedasticity simultaneously with the estimation of the functional form of the systematic part and the usefulness of using a flexible form for this correction;

iii) established that taking autocorrelation, or heteroscedasticity, or both, into account could change the functional form of the systematic part of the model and that freeing the functional form could modify the autocorrelation structure; finally, we found that taking heteroscedasticity into account could also modify the autocorrelation structure;

v) found that elasticities of transit and gasoline demand and implicit values of transit time were quite sensitive to the specification of the stochastic part of the model.

In our Appendix, we analyse further the problem of lack of invariance of the asymptotic t-tests of the β_k coefficients to unit of measurement specifications of the X_k when Box-Cox transformations are used. Our example illustrates both the considerable difference between conditional and unconditional t-values and the relatively small effect that scaling with arithmetic averages would have had on our results.

APPENDIX

ON THE t-STATISTICS OF THE REGRESSION COEFFICIENTS

In this Appendix, we illustrate some of the points made in Section 3.3 concerning possible solutions to the problem that, when BCT are used, as in (12-A) and (12-B), the associated regression coefficients (the β_k and the δ_m) have asymptotic t-statistics which are not invariant to the units of measurement of the X_k and Z_m. Our illustration is based on the S equation (with the distribution of residuals assumed to be spherical), and the results are reported in Table 6. In the first four columns, the scaling rule of the L-1.1 algorithm is used : y and the X_k are scaled by the power of 10 which yields a series with a mean of the order of magnitude of 1. The last column to the right (2.0R) uses the arithmetic mean of the regressand and regressors to scale them, a procedure which yields coefficients equal to the intuitive sample elasticities $\eta(S)$ defined in (34). It is clear that the unconditional t-values obtained under these two scaling rules are, in general, very close, as a comparison of columns 2.0 and 2.0R shows. Moreover, the first three columns present the t-values conditional upon the value of λ (applied to the dependent variable and variables 15-26), with $\lambda = 0$, $\lambda = 1$ and $\lambda = \hat{\lambda}$ respectivaly used in experiments 0.0, 1.0 and 2.0C. These conditional values are invariant to the scaling specification. The comparison of the conditional values in column 2.0C and the unconditional values in column 2.0 shows in our case significant differences.

TABLE 6

CONDITIONAL AND UNCONDITIONAL t-STATISTICS* WITH DIFFERENT SCALING RULES**

Schoolchildren equation from Table 1 : $t(\hat{\beta}_k)$

Condition		(λ=0)	(λ=1)	($\lambda=\hat{\lambda}$)		
Experiment		0.0	1.0	2.0C	2.0	2.0R
X_k : 1	STR	-13,94	-11.51	-12.74	-3.49	-4.77
2	STRUCT1	1.21	.35	.69	.43	.44
3	TLK	2.28	3.17	2.71	1.61	1.80
4	RFK	-1.25	-.86	-1.07	-.79	-.78
5	SFK	-2.43	-3.25	-2.94	-1.95	-2.36
6	CSFMR	1.52	1.79	1.58	1.13	1.21
7	RSCMP	12.59	11.84	12.21	4.41	8.44
8	CHEAT2	-2.20	-1.16	-1.34	-1.29	-1.26
9	CARDEX	.14	.40	.28	.19	.19
10	EX7	5.93	5.38	5.74	3.45	6.02
11	EX8	2.00	1.41	1.72	1.31	1.41
12	ZSC	2.82	3.56	3.28	1.43	1.69
13	ANT	2.19	4.13	3.13	.00	.00
14	ANTL1	-1.16	-1.24	-1.32	-.00	-.00
15	FSCD	-1.29	-2.30	-1.78	-1.03	-1.10
16	P	-1.08	-.70	-1.07	-.73	-.73
17	NC	.20	-1.37	-1.11	-.77	-.77
18	TW	-.84	-.32	-1.35	-.85	-.86
19	TT	1.63	.80	1.28	.68	.68
20	T2T	1.28	1.27	1.79	1.23	1.33
21	RAWEM	-.28	1.49	.33	.26	.26
22	SHONF	2.40	3.73	2.96	2.59	2.68
23	TEND	1.15	2.18	2.11	1.72	1.76
24	WD	3.34	3.88	3.70	2.42	2.62
25	SAT	1.30	1.50	1.62	1.44	1.41
26	SHD	1.01	.94	1.04	1.02	1.02
27	CONSTANT	-1.78	-3.07	-2.74	-1.99	-6.58

* All values cut after second decimal without rounding.

** In the first 4 columns, the scaling of the L-1.1 algorithm is used; in column 2.0R, all variables are scaled by their arithmetic means.

ACKNOWLEDGEMENTS

The research for this paper was supported at various stages by Transport Canada, the F.C.A.C. program of the ministère de l'Education du Québec, the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council of Canada. A first and widely circulated version, "Multiple Regression Analysis with Box-Cox Transformations : a Transportation Application", C.R.T., Publication # 166 (1980), contained partial results on one of the three equations analysed here. Early estimates were presented at the June 1979 meeting of the Econometric Society in Montreal. Lina and Lisa DiPietro made significant contributions as research assistants. A part of Gaudry's contribution was made during his tenure as an Alexander von Humboldt Research Scholar in 1985 at Karlsruhe Universität where Reiner Koblo provided computing assistance.

REFERENCES

1. Amemiya, T., 1973, Regression Analysis when the Dependent Variable is Truncated Normal, Econometrica 41, 4, 997-1016.

2. Amemiya, T. and J.L. Powell, 1980, A Comparison of the Box-Cox Maximum Likelihood Estimator and the Nonlinear Two Stage Least Squares Estimator, Technical report no. 322, Stanford University.

3. Appelbaum, E., 1979, On the Choice of Functional Forms, International Economic Review 20, 2, 449-458.

4. Beach, C.M. and J.G. MacKinnon, 1978, A Maximum Likelihood Procedure for Regression with Autocorrelated Errors, Econometrica 46, 1, January, 51-58.

5. Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation and Inference in Nonlinear Structural Models, Annals of Economic and Social Measurement 3, 653-665.

6. Berndt, E.K. and M.S. Khaled, 1979, Parametric Productivity Measurement and Choice Among Flexible Functional Form, Journal of Political Economy 87, 6, 1220-1245.

7. Bickell, P.J. and K.A. Doksum, 1981, An Analysis of Transformations Revisited, Journal of the American Statistical Association 76, no. 374, 296-311.

8. Blackley, P., J.R. Follain and J. Ondrich, 1984, Box-Cox Estimation of Hedonic Models : How Serious is the Iterative OLS Variance Bias ? The Review of Economics and Statistics 66, 348-353.

9. Blaylock, J.R. and D.M. Smallwood, 1983, Interpreting the Effect of Binary Variables in Transformed Models, Economics Letters 12, 3-4, 255-259.

10. Box, G.E.P. and D.R. Cox, 1964, An Analysis of Transformations, Journal of the Royal Statistical Society B, 211-243.

11. Box, G.E.P. and G.M. Jenkins, 1970, Time Series Analysis Forecasting and Control (Holden-Day, San Francisco).

12. Burguette, J.F., A.R. Gallant and G. Souza, 1982, On Unification of the Asymptotic Theory of Nonlinear Econometric Models, Econometric Reviews, 1, 2, 151-190.

13. Buse, A., 1979, A Note on Generalized Heteroskedasticity, working paper, Economics Department, University of Alberta.

14. Carroll, R.J. and D. Ruppert, 1981, On Prediction and the Power Transformation Family, Biometrika 68, no. 3, 609-615.

15. Dagenais, M.G., 1982, The Tobit Model with Serial Correlation, Economics Letters 10, 3-4, 263-67.

16. Dagenais, M.G., M.J.I. Gaudry and T.C. Liem, 1980, Multiple Regression Analysis with Box-Cox Transformations and Nonspherical Residual Errors : a Transportation Application, publication no. 166, Centre de Recherche sur les Transports, Université de Montréal.

17. Dagum, C. and E.B. Dagum, 1974, Construction de modèles et analyse économétrique, Cahiers de l'Institut des Sciences Mathématiques et Economiques Appliquées 8, no. 11-12.

18. Draper, N.R. and D.R. Cox, 1969, On Distributions and Their Transformations to Normality, Journal of the Royal Statistical Society B 31, 472,476.

19. Egy, D. and K. Lahiri, 1979, On Maximum Likelihood Estimation of Functional Form and Heteroskedasticity, Economics Letters 2, 2, 155-160.

20. Fletcher, P. and M.J.D. Powell, 1963, A Rapidly Convergent Descent Method for Minimization, Computer Journal 6, 163-168.

21. Gaudry, M., 1980, A Study of Aggregate Bi-Modal Urban Travel Supply, Demand and Network Behavior Using Simultaneous Equations with Autoregressive Residuals, Transportation Research B 14, 1/2, 29-58.

22. Gaudry, M., 1978, Seemingly Unrelated Static and Dynamic Urban Travel Demands, Transportation Research 12, 3, 195-211.

23. Gaudry, M., 1975, An Aggregate Time-Series Analysis of Urban Transit Demand : the Montreal Case, Transportation Research 9, 249-258.

24. Gaudry, M.J.I. and M.G. Dagenais, 1978, The Use of Box-Cox Transformations in Regression Models with Heteroskedastic Autoregressive Residuals, publication # 106, Centre de recherche sur les transports, Université de Montréal.

25. Gaudry, M.J.I. and M.G. Dagenais, 1979, Heteroscedasticity and the Use of Box-Cox Transformations, Economics Letters 2, 3, 225-229.

26. Gaudry, J.J.I. and M.J. Wills, 1978, Estimating the Functional Form of Travel Demand Models, Transportation Research 12, 257-289.

27. Glejser, H., 1969, A New Test for Heteroskedasticity, Journal of the American Statistical Association 64, 316-323.

28. Goldberger, A.S., 1968, The Interpretation and Estimation of Cobb-Douglas Functions, Econometrica 35, 3-4, 464-472.

29. Greene, D., 1982, State-level Stock System Model of Gasoline Demand, Transportation Research Record 801, 44-50.

30. Halvorsen, R. and R. Palmquist, 1980, The Interpretation of Dummy Variables in Semilogarithmic Equations, American Economic Review 70, 3, 475-475.

31. Heckman, J. and S. Polachek, 1974, Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship, Journal of the American Statistical Association 69, 350-354.

32. Huang, C.J. and O.R. Grawe, 1980, Functional Forms and the Demand for Meat in the United States : A Comment, Review of Economics and Statistics 62, 1, 144-146.

33. Kau, J.B. and C.F. Sirmans, 1976, The Functional Form of the Gravity Model : A New Technique with Empirical Results, working paper, department of Finance, University of Illinois. Presented at the North American Meeting of the Regional Science Association in Toronto.

34. Kmenta, J., 1971, Elements of Econometrics (McMilland, New York), p. 461.

35. Liem, T.C., M.G. Dagenais and M.J.I. Gaudry, 1983, L-1.1 : A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals, publication # 301, Centre de recherche sur les transports, Université de Montréal.

36. Olsen, R.J., 1978, The Method of Box and Cox, A Pitfall, working paper # 804, Institution for Social and Policy Studies, Yale University.

37. Park, R.E., 1966, Estimation with Heteroskedastic Error Terms, Econometrica 34, 888.

38. Poirier, D.J., 1978, The Use of the Box-Cox Transformation in Limited Dependent Variable Models, Journal of the American Statistical Association 73, 284-287.

39. Rao, C.R., 1948, Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Application to Problems of Estimation, Proceedings of the Cambridge Philosophical Society, 44, 50-57.

40. Robinson, P.M., 1982, On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables, Econometrica 50, 27-41.

41. Rosett, R.N. and F.D. Nelson, 1975, Estimation of the Two-limit Probit Regression Model, Econometrica 43, no. 1, 141-146.

42. Savin,N.E. and K.J. White, 1978, Estimation and Testing for Functional Form and Autocorrelation : A Simultaneous Approach, Journal of Econometrics 8, 1-12.

43. Schlesselmann, J., 1971, Power Families : A Note on the Box and Cox Transformation, Journal of the Royal Statistical Society B 33, 307-311.

44. Seaks, T.G. and S.K. Layson, 1983, Box-Cox Estimation with Standard Econometric Problems, The Review of Economics and Statistics, 65, 1, 160-164.

45. Spitzer, J.J., 1976, The Demand for Money, the Liquidity Trap, and Functional Forms, International Economic Review 17, 220-227.

46. Spitzer, J.J., 1977, A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms, Journal of Economics 5, 117-128.

47. Spitzer, J.J., 1982, A Primer on Box-Cox Estimation", The Review of Economics and Statistics 64, 307-313.

48. Spitzer, J.J., 1984, Variance Estimates in Models with the Box-Cox Transformation : Implications for Estimation and Hypothesis Testing, The Review of Economics and Statistics 66, 645-652.

49. Theil, H., 1971, Principles of Econometrics (John Wiley & Sons Inc., New York).

50. Tobin, J., 1958, Estimation of Relationships for Limited Dependent Variables, Econometrica 25, 24-36.

51. Wang, G.H.K., W. Maling and E. McCarthy, 1981, Functional Forms and Aggregate U.S. Domestic Air Cargo Demand : 1950-1977, Transportation Research A15, 249-256.

52. Wang, G.H.K. and D. Skinner, 1984, The Impact of Fare and Gasoline Price Changes on Monthly Transit Ridership : Empirical Evidence from Seven U.S. Transit Authorities, Transportation Research B18, 29-41.

53. Welland, J.D., 1978, Cognitive Abilities, Schooling and Earnings : the Question of Functional Form, The Review of Economics and Statistics, 60, 622-627.

54. White, K.J. and F. Ho, 1982, UBC SHAZAM - An Econometrics Computer Program, Version 4.5, The University of British Columbia Computing Centre, November.

55. Zarembka, P., 1968, Functional Form in the Demand for Money, Journal of the American Statistical Association 63, 502-511.

56. Zarembka, P., 1974, Transformation of Variables in Econometrics, in P. Zarembka, ed., Frontiers in Econometrics (Academic Press, New York), ch. 3.

57. Zellner, A., 1971, An Introduction to Bayesian Inference in Econometrics, (John Wiley and Sons, Inc., New York).

FOOTNOTES

[1]. The BCT estimates were 0.92 for the adult market and 0.72 for the school-children market. These values are taken from below in Table 1, Experiment # 2.1, Series A-1 and S-1; they differ slightly from the less precise estimates shown on Figures 9 and 15 of the Gaudry and Wills (1978) paper.

[2]. Note that, when the number of observations is small, it might be advantageous to recuperate the first r observations, as pointed out by Beach and MacKinnon (1978). Recuperating the first r observations does not raise any theoretical problem, but the expressions of the log-likelihood function and of its first derivatives then become more messy: we have therefore not considered this alternative.

[3]. Strictly speaking, they are conditional maximum likelihood estimators, conditional on the values of the first r observations. Similarly, when we speak of the probability that $y_t^{(\lambda)}$ be equal to its upper or lower limit, or when we discuss the calculation of $E(y_t)$, these are all *conditional* on the past values of y.

[4]. This point has also been recognized by other authors in the literature, for the original Box-Cox model. For example, Zellner (1971, p. 163, fn. 2) notes that the range of $y_t^{(\lambda)}$ is not $-\infty$ to $+\infty$ "but just a subinterval of this range". He then points out that if the probability that $y_t^{(\lambda)}$ "will fall in the excluded interval is small, the normality assumption will not be vitally affected". In the same line of thought, it is readily seen in Amemiya and Powell (1980, p. 6, equation 2.10) that, if it is assumed that $y_t^{(\lambda)}$ has a truncated normal distribution and if the surface under the normal curve for the admissible region is close to one, the asymptotic bias of the "Box-Cox MLE" of λ vanishes; in their notation, if

$$\Phi\left(\frac{1}{\lambda_o \sigma_o} + \frac{\mu_{to}}{\sigma_o}\right) \to 1 \text{ for every t , then } -\frac{1}{\sigma_o \lambda_o^2} \sum \frac{\Phi\left(\frac{1}{\lambda_o \sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)}{\Phi\left(\frac{1}{\lambda_o \sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)} \text{ will } \to 0 \ .$$

Clearly, a rigorous analytical derivation of the asymptotic properties of the proposed estimators of the model discussed in the present paper would require further investigation. An expensive alternative approach, which could generate information on the finite sample properties as well as on the asymptotic properties of the estimators, would be to resort to Monte Carlo experiments.

[5]. In Amemiya's (1973) paper, the proofs that are developed seem perfectly compatible with the fact that a given *finite* observable sample may contain zero limit observations. The difficulty of extending Amemiya's proof to the model discussed in the present paper comes mainly from the assumption of serial correlation. The introduction of serial correlation in limited dependent variable models complicates greatly the general expression of the likelihood function, since every run of m limit observations leads to the introduction of an m-dimensional integral of the multivariate normal density function [Dagenais (1982)]. However, Robinson's approach (1982) could possibly be used in this case.

[6]. In certain cases, one may not be ready to make such an assumption and might prefer to specify the value of L explicitly. In such case, the expression for $E(y_t)$ would become:

$$\text{for } \lambda_y > 0 \quad : \quad E(y_t) = \int_{w_t^{**}}^{w_t^*} y(w)_t \; f(w;\sigma^2)\; dw + L \int_{w_t^*}^{\infty} f(w;\sigma^2)\; dw$$

and

$$\text{for } \lambda_y = 0 \quad : \quad E(y_t) = \int_{-\infty}^{\bar{w}_t} \bar{y}(w)_t \; f(w;\sigma^2)\; dw + L \int_{\bar{w}_t}^{\infty} f(w;\sigma^2)\; dw$$

where

$$\bar{w}_t = \frac{\ln L}{\sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \; y^*_{t-\ell} - \sum_{\ell=1}^{K} \beta_k \; X^{**}_{kt} \quad .$$

[7]. Because of these considerations, in our program [Liem et al. (1983)], the user is asked to specify the value of L. Otherwise, if the user *does not* specify L and the maximum likelihood estimate of λ_y turns out to be *negative*, the program generates, by default, its own approximation of L and uses it to compute $E(y_t)$ from equation (22). When the L value is thus generated by the program, it is automatically printed in the output, so that the user may be made aware of the approximation actually used. This procedure is further discussed in Liem et al. (1983).

[8]. If that was not the case, it was usually easy to change the AR structure in such a way as to obtain the desired result, by trial and error. The reader should also note our Section 4.3.3 on this point.

[9]. The fact that the norm defined for y (namely its sample arithmetic mean $\bar{y}$ or its sample geometric mean $\dot{y}$) is itself a stochastic variable also raises statistical complications which have not been considered, for example, by Spitzer (1984).

[10]. Note however that likelihood ratio tests and Rao-type tests [Rao (1948)] are invariant with respect to changes in the units of measurement. Performing such tests for each individual β parameter would however be much too costly, in our case.

[11]. It is unclear how this unrolling is performed in UBC SHAZAM [White and Ho, (1982), p. 58].

[12]. This was case 5 in the previous version of this paper (Dagenais et al., 1980).

[13]. Note, however, that *estimated* elasticities are evidently stochastic measures.

[14]. Note that our L-1.1 algorithm does not provide for the possibility that quasi-dummy variables be affected by a BCT (with $\lambda \neq 0$). This is not because such a transformation would not be invariant to units of measurement —one could compensate for the effect on zero observations with a binary variable.

[15]. Clearly, base levels for the dependent variable and explanatory variables are arbitrary and could all be different.

[16]. Note that the vector $\bar{\bar{X}}$ might contain other quasi-dummy variables besides X_k. When interest centers exclusively on the effect of X_k when the corresponding event occurs, the other dummy variables may be ascribed their usual $\bar{X}$ values.

[17]. Formally, for the model defined in (12), the expected value elasticity, written for any observation t, is

$$\eta_{kt}(E) = \frac{1}{E(y_t)} \int_{R(w)} [y_t(w)]^{1-\lambda_y} [X_{kt}^{\lambda_{xk}} \beta_k + H_{kt}(w)] N(w)\, dw$$

where $N(w) = f(w;\sigma^2)$ in (18), $E(y_t)$ and $y_t(w)$ are defined in (18), $R(w)$ is the domain of integration for w (see Liem et al. (1983) for details) and $H_{kt}(w)$ equals 0 if X_k is not a Z_m variable and otherwise equals

$$\frac{1}{2} Z_{mt}^{\lambda_{zm}} \delta_m \sqrt{f(Z_t)} \left[\sum_\ell \rho_\ell \left(y^*_{t-\ell} - \sum_k X^*_{k,t-\ell} \beta_k\right) + w\right], \text{ where } X^*_{kt} = \frac{X_{kt}^{(\lambda_{xk})}}{\sqrt{f(Z_t)}}.$$

[18]. In effect, the numerator is equal to

$$\int_{R(w)} y(w_t)_{X=\tilde{X}_1} f(w;\sigma^2)\, dw - \int_{R(w)} y(w_t)_{X=\tilde{X}_o} f(w;\sigma^2)\, dw$$

where the domain of integration is $[w^{**}_t, \infty]$ for $\lambda_y > 0$, $[-\infty, +\infty]$ for $\lambda_y = 0$ and $[-\infty, w^*_t]$ for $\lambda_y < 0$.

[19]. Clearly, this case is the analog of (38). If we had used $E(y|\bar{\bar{X}})$ in (38), then (42-3) would be the corresponding expression.

[20]. If that were not true, the expressions for the linear case would be more complicated. For instance, case 1.A in Table 2 would be :

$$\frac{\beta_d \int_{w_t^{**}}^{\infty} f(w;\sigma_w^2)\, dw}{\int_{w_t^{**}}^{\infty} y(w_t)\, f(w;\sigma_w^2)\, dw} .$$

[21]. For example, in the simpler intuitive case, the partial derivative $(\partial y/\partial X_k) = (\beta_k / y^{\lambda_y - 1})$ can itself be derived with respect to X_k to verify this point. The result, $\partial(\partial y/\partial X_k)/\partial X_k = -\beta_k^2 (\lambda_y - 1)/y^{(2\lambda_y - 1)}$, has a denominator of the order of 1 because of our scaling of y described in the Appendix and a numerator which is small, for small values of $|\lambda_y - 1|$, if $|\beta_k| < 1$. In Table 3, $|\lambda_y - 1|$ as well as the coefficients of variables STR, FIR, EX7, EX8, EX9, EX0, EX1, ZA and ZSC are all smaller than 1 in absolute value except that of ZSC which is of the order of 1.5.

[22]. The numerators of the two expressions differ. In (43), we have $\beta_k \int_{R(w)} [y_t(w)]^{1-\lambda_y} N(w)\, dw$; in (42-1), the numerator is $\int_{R(w)} [y_t(w)_1]\, N(w)\, dw - \int_{R(w)} [y_t(w)_o]\, N(w)\, dw$. In the intuitive case, the values for $\lambda_y = 0$ are respectively β_o and $e^{\beta_d} - 1$. Considering that the first two terms of the Taylor expansion of e^{β_d} are $1 + \beta_d$, the approximation is acceptable if $|\beta_d| < 1$. In Table 3, the coefficients of the dummy variables STRUCT1, CHEAT2 and CARDEX, which have a positive value for several observations (ANT and ANTL1 are defined only for one observation each), are between 0.02 and 0.18.

[23]. We have pointed out in Section 3.3 that the t-values associated with the coefficients of X_k and Z_m were not invariant to units of measurement when a BCT was estimated. See also the Appendix on this point.

[24]. We do not know whether university attendance reduces other social activities of students sufficiently to imply a net gain or loss.

[25]. It is known that the integration of bus and subway lines was made to the benefit of adults; schoolchildren cannot have gained by the increased access times implied by the integrated network.